DOI:
10.1016/j.neunet.2011.10.002

Document Version
Peer reviewed version

Neural Networks, 26:7-28, 2012.

Predictive coding as a model of the V1 saliency map hypothesis

M. W. Spratling
King's College London, Department of Informatics and Division of Engineering, London. UK.

Abstract

The predictive coding/biased competition (PC/BC) model is a specific implementation of predictive coding theory that has previously been shown to provide a detailed account of the response properties of orientation tuned cells in primary visual cortex (V1). Here it is shown that the same model can successfully simulate psychophysical data relating to the saliency of unique items in search arrays, of contours embedded in random texture, and of borders between textured regions. This model thus provides a possible implementation of the hypothesis that V1 generates a bottom-up saliency map. However, PC/BC is very different from previous models of visual salience, in that it proposes that saliency results from the failure of an internal model of simple elementary image components to accurately predict the visual input. Saliency can therefore be interpreted as a mechanism by which prediction errors attract attention in an attempt to improve the accuracy of the brain's internal representation of the world.

Keywords: saliency map; texture segmentation; visual search; contour integration; neural network model; attention

1 Introduction

A number of psychophysical experiments suggest that primary visual cortex (V1) may be involved in the computation of visual salience (Koene & Zhaoping, 2007; Zhaoping, 2008; Zhaoping et al., 2009a; Zhaoping & May, 2007; Zhaoping et al., 2009b). These experiments thus support the hypothesis that V1 operates as a bottom-up, pre-attentive, saliency map (Li, 2002). Previous work (Spratling, 2010, 2011) has demonstrated that a simple functional model (PC/BC), derived from the predictive coding and biased-competition theories of cortical function (Spratling, 2008a,b), can simulate a very wide range of V1 response properties including orientation tuning, size tuning, spatial frequency tuning, temporal frequency tuning, cross-orientation suppression, and surround suppression. This article extends that work by showing that the PC/BC model of V1 can also simulate a wide range of psychophysical experiments on visual salience, and hence, demonstrates that PC/BC provides a possible implementation of the V1 saliency map hypothesis.

Predictive coding is a scheme for combining bottom-up evidence with prior knowledge to infer the most likely causes of a sensory stimulus (Bubic et al., 2010; Rao & Ballard, 1999). This is achieved through an iterative process in which a prediction about the underlying causes of the sensory data (*e.g.*, an internal representation of the world), is used to reconstruct the expected sensory input. These predicted inputs are compared with the actual stimulus-driven activity in order to calculate the residual error between the predicted data and the sensory evidence. This error is then used to modify the predicted causes to form a more accurate internal model of the world, which will in turn reduce the residual error. Predictive coding is a specific example of more general theories of efficient encoding or redundancy reduction (Attneave, 1954; Barlow, 2001; Olshausen & Field, 1996b, 1997), of generative models of inference and learning (Hinton, 2002; Hinton et al., 1995; Hoyer, 2004; Hoyer & Hyvärinen, 2000; Lee & Seung, 1999; Olshausen & Field, 1996a), and of theories of hierarchical perceptual inference or analysis-by-synthesis (Barlow, 1994; Friston, 2005; Lee & Mumford, 2003; Mumford, 1992; Yuille & Kersten, 2006). See Spratling (in press) for a more in-depth discussion of the relationship between predictive coding and other theories of cortical function.

Predictive coding is also a particular instantiation of the free-energy principle (Friston, 2010, 2009). Free-energy suggests that sensory prediction errors give rise to action that will reduce this error (Friston, 2010; Friston et al., 2010). Hence, if perceptual salience has a role in the control of action (*e.g.*, in directing eye movements or in the allocation of endogenous attention) then saliency should be correlated with the prediction errors generated in a predictive coding model. In order to test this hypothesis, measurements were made of the residual errors generated by the PC/BC model. For a very wide range of images, the relative strength of the error calculated by PC/BC at different locations in the image was found to be consistent with the perceptual saliency of those different parts of the image. The model was tested by comparing the saliency values calculated from the residual error generated by the PC/BC model with psychophysical measures of saliency (*i.e.*, reaction times and response accuracy) recorded in experiments on texture segmentation, visual search, and contour integration. The model was found to provide an accurate account of perceptual saliency in all of these domains.

Specifically, in tasks evaluating the saliency of the border between textured regions, the model is shown to account for: the effects of orientation contrast (section 3.1); the effects of element spacing (section 3.2); and of superimposed irrelevant texture elements (section 3.3). In tasks evaluating the saliency of unique elements in search arrays, the model is shown to account for the range of search efficiencies ('pop-out', 'serial', and 'parallel' search) found in psychophysical experiments (section 3.4); asymmetries in visual search (section 3.5); the effects of element spacing (section 3.6); the effects of superimposed irrelevant texture elements (section 3.7); the effects of abrupt element onsets (section 3.8); to account for the preview effect (section 3.9); and to account for 'flicker' induced change blindness (section 3.10). The model is also used to explore the possible effects on saliency of cortical feedback generated by expectation, attention, or object familiarity. In these experiments, the model is shown to account for the saliency of a contour embedded in random texture elements (section 4.1); the effects of element novelty and familiarity in feature search (section 4.2); the effects of prior exposure to features of either the target or distractors in a subsequent search array, *i.e.*, "priming of pop-out" and the "distractor previewing effect" (section 4.3); the preview effect (section 4.4); the increase in target saliency when distractor elements form a familiar contour (section 4.5); the effects of contextual guidance in feature search (section 4.6); and the saliency of objects that are incongruous with the scene (section 4.7). The results of these experiments also suggest that predictive coding can provide a natural explanation for the faster recognition of objects congruent with a visual scene despite the fact that incongruent objects are more likely to attract attention (section 4.7), and inhibition-of-return (section 4.8).

For all these seemingly diverse experiments on visual saliency, the model provides a single computational explanation of the results. The model proposes that V1 encodes visual information using an over-complete set of Gabor basis functions, which provided a means of accurately and efficiently representing natural images (Field, 1987, 1994; Olshausen & Field, 1997, 2005). Salient locations in an image are those locations where this representation is least accurate.

2 Model Description

2.1 The Retina/LGN Model

To simulate the effects of circular-symmetric center-surround receptive fields (RFs) in LGN and retina, the input to the PC/BC model of V1, described below, was an input image (I) pre-processed by convolution with a Laplacian-of-Gaussian (LoG) filter (l) with a standard deviation equal to 1.5 pixels, The output from this filter was subject to a multiplicative gain (the strength of which was determined by parameter κ) followed by a saturating non-linearity, such that:

$$\mathbf{X} = \tanh\left\{\kappa(I * l)\right\} \tag{1}$$

A value of $\kappa = 2\pi$ was used in all experiments reported here.

The positive and rectified negative responses were separated into two images $\mathbf{X}_{ON}$ and $\mathbf{X}_{OFF}$ simulating the outputs of cells in retina and LGN with on-center/off-surround and off-center/on-surround RFs respectively. These on- and off-channels provided the input to the PC/BC model of V1. This pre-processing stage described above is illustrated on the left of Figure 1.

2.2 The V1 Model

The PC/BC model of V1 is illustrated on the right of Figure 1 and described by the following equations:

$$\mathbf{E}_o = \mathbf{X}_o \oslash \left(\epsilon_2 + \sum_{k=1}^{p} \left(\hat{w}_{ok} * \mathbf{Y}_k\right)\right) \tag{2}$$

$$\mathbf{Y}_k \leftarrow \left(\epsilon_1 + \mathbf{Y}_k\right) \otimes \sum_{o} \left(w_{ok} \star \mathbf{E}_o\right) \tag{3}$$

$$\mathbf{Y}_k \leftarrow \mathbf{Y}_k \otimes \left(1 + \eta \mathbf{A}_k\right) \tag{4}$$

Where $o \in [ON, OFF]$; $\mathbf{X}_o$ is a 2-dimensional array, equal in size to the input image, that represents the input to the model of V1; $\mathbf{E}_o$ is a 2-dimensional array, equal in size to the input image, that represents the error-detecting neuron responses; $\mathbf{Y}_k$ is a 2-dimensional array, equal in size to the input image, that represents the prediction neuron responses; $\mathbf{A}_k$ is a 2-dimensional array, equal in size to the input image, that represents the weighted sum of top-down predictions arising from extrastriate cortical regions not explicitly modeled here; w_{ok} is a 2-dimensional kernel representing the synaptic weights for a particular class (k) of neuron normalized so that the

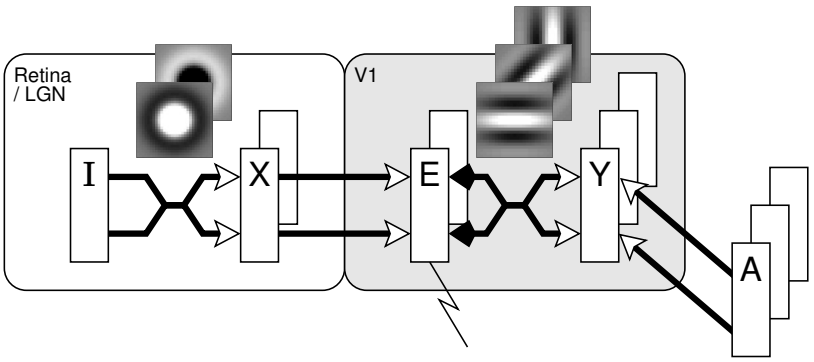

Figure 1: The PC/BC model of V1 (right) and the model of retina/LGN (left). The input image I was preprocessed by convolution with a circular-symmetric on-center/off-surround kernel to generate the input to the ON channel of the V1 model, and a circular-symmetric off-center/on-surround kernel to generate the input to the OFF channel of the V1 model. The prediction neurons, labeled $\mathbf{Y}$, represent V1 simple cells. The activity of these neurons was simulated by convolving the outputs of the ON and OFF channels of the error-detecting neurons, labeled $\mathbf{E}$, with the ON and OFF channels of a number of weight kernels (defined by Gabor functions) representing V1 RFs. This convolution process effectively reproduces the same RFs at every pixel location in the image. The prediction neuron responses could also be modulated by feedback from higher cortical regions which were not explicitly modeled, rather these effects were simulated by additional inputs to the V1 model, labeled $\mathbf{A}$. The responses of the error-detecting neurons were influenced by divisive feedback from the prediction neurons, which was also calculated by convolving the prediction neuron outputs with the weight kernels. Responses of the error-detecting neurons were recorded during experiments and the strength of response at each location was assumed to relate to the saliency of that location in the image.

sum of all the weights is equal to ψ; $\hat{w}_{ok}$ is a 2-dimensional kernel representing the same synaptic weights as w_{ok} but normalized so that the maximum value is equal to ψ; p is the total number of kernels; ϵ_1, ϵ_2, η and ψ are parameters; $\oslash$ and $\otimes$ indicate element-wise division and multiplication respectively; $\star$ represents cross-correlation (which is equivalent to convolution without the kernel being rotated 180^o); and $*$ represents convolution (which is equivalent to cross-correlation with a kernel rotated by 180^o). Parameter values $\psi = 5000$, $\epsilon_1 = 0.0001$, $\epsilon_2 = 250$, and $\eta = 1$ were used in the simulations reported in this article. These equations were evaluated in the order 2, 3, 4 and the values of $\mathbf{Y}_k$ given by equation 4 were then substituted back into equations 2 and 3 to recursively calculate the changing neural activations at each time-step.

Equation 2 describes the calculation of the neural activity for each population of error-detecting neurons. These values are a function of the activity of the input to V1 divisively modulated by a weighted sum of the outputs of the prediction neurons in V1. The activation of the error-detecting neurons can be interpreted as representing the residual error between the input and the reconstruction of the input generated by the prediction neurons. The values of $\mathbf{E}$ indicate the degree of mismatch between the reconstruction of the input and the actual input (assuming ϵ_2 is sufficiently small to be negligible[1]). When a value within $\mathbf{E}$ is greater than $\frac{1}{\psi}$ it indicates that a particular element of the input is under-represented in the reconstruction, a value of less than $\frac{1}{\psi}$ indicates that a particular element of the input is over-represented in the reconstruction, and a value of $\frac{1}{\psi}$ indicates that the reconstruction perfectly predicts the bottom-up stimulation. In the simulations reported here, the activation of each error-detecting neuron was assumed to be related to the saliency at the corresponding image location, with high error values (indicating parts of the stimulus that are under-represented by the prediction neurons) corresponding to high saliency.

Equation 3 describes the updating of the prediction neuron activations. The values of $\mathbf{Y}_k$ represent predictions of the causes underlying the inputs to the model of V1. The response of each prediction neuron is a function of its activation at the previous iteration and a weighted sum of afferent inputs from the error-detecting neurons. The strength of this feedforward input will be influenced by the strength of the error neuron outputs (which are in turn influenced by the sensory-driven input and the accuracy of the predictions, as described in the preceding paragraph) and by the similarity between this activity and the synaptic weights. If the input remains constant, the values of $\mathbf{Y}_k$ will converge over multiple iterations to steady-state values that reconstruct the input with minimum error (Spratling, in press), and hence, this pattern of activations constitutes the best estimate of the causes of the

[1] Although $\epsilon_2 = 250$ seems large, it is negligible in comparison to typical values calculated by $\sum_{k=1}^{p} (\hat{w}_{ok} * \mathbf{Y}_k)$, which is a sum of many products calculated (due to the large value of parameter ψ) using values of $\hat{w}_{ok}$ that are large.

input, given the prior knowledge stored in the synaptic weight values. Whether or not a particular prediction neuron is included in the set of active neurons forming the best estimate of the causes of the input, will depend not only on how well that neuron's RF matches features of the input stimulus, but also on how well other prediction neurons can represent the same features of the stimulus. Hence, the iterative procedure to determine the prediction neuron responses can also be interpreted as a mechanism of competition between the prediction neurons, in which prediction neurons with overlapping RFs compete for the right to represent the features of the stimulus falling within their RFs.

Equation 4 describes the effects of cortical feedback connections, as described in the next section.

2.3 Modeling the Effects of Cortical Feedback

PC/BC is a hierarchical model, and hence, it proposes that the V1 model is one part of a larger hierarchy of processing stages. Equation 4 describes the effects on the V1 prediction neuron activations of top-down inputs from prediction neurons at later processing stages (*i.e.*, in extrastriate cortical regions). These top-down inputs have a purely modulatory effect that enables predictions generated by neurons higher up a processing hierarchy to influence the strength of each prediction made at the current processing stage. Rather than attempting to simulate a large number of interacting extrastriate cortical regions, the effects of cortical feedback are more conveniently and controllably modeled by using an array of inputs ($\mathbf{A}$) to the V1 model which represent the sum of the top-down predictions generated at the subsequent processing stages. There is one element of $\mathbf{A}$ for each element of $\mathbf{Y}$, and hence, one top-down input to each V1 prediction neuron. This makes it possible to specify exactly which prediction neurons will receive feedback. In the simulations reported in section 4, feedback is selectively applied to neurons with particular orientation preferences, to neurons with RFs within a particular range of spatial locations, or neurons with a particular combination of both location and orientation preference. For all prediction neurons receiving top-down feedback, the corresponding element of $\mathbf{A}$ was set to a value of 0.25, and all other elements of $\mathbf{A}$ were set to zero. For all simulations reported in section 3, top-down predictions from extrastriate regions were assumed to be negligible, and all elements of $\mathbf{A}$ were given a value of zero, in which case equation 4 has no effect.

Cortical feedback might be generated by numerous different cortical regions for different purposes. For example, it might be generated by visually-driven neurons with larger receptive fields when the stimulus forms some familiar shape. It might alternatively be driven by endogenous attention mechanisms. While these different forms of feedback are clearly not functionally equivalent, they are treated identically in the current model. Whatever the source, top-down activity will have two effects on the PC/BC model of V1. (1) Increasing the response of the prediction neurons that represent information consistent with the top-down expectation (see equation 4). This will result in these prediction neurons sending stronger feedforward activation, and hence, make this information more conspicuous for cortical regions at subsequent stages along the processing hierarchy. (2) The enhanced activity in the prediction neurons consistent with top-down expectations will in turn decrease the response of the error-detecting neurons from which these prediction neurons receive their input (see equation 2). Since the strength of the responses of the error-detecting neurons are assumed to be related to saliency, this will result in the salience of the corresponding location being reduced, and hence, making those locations less likely to capture attention.

2.4 Defining the RFs of V1 Model Neurons

The RF of a simple cell in primary visual cortex can be accurately modeled by a 2-dimensional Gabor function (Daugman, 1980, 1988; Jones & Palmer, 1987; Lee, 1996; Marcelja, 1980). Hence, the Gabor function was used to define the weights of each kernel w_{ok}. A definition of a Gabor function of the form proposed by Lee (1996) was used, which includes a term to remove the D.C. response of the filter:

$$g(\sigma, \gamma, \lambda, \phi, \theta) = \exp\left\{-\frac{x'^2 + (y'/\gamma)^2}{2\sigma^2}\right\}\left[\cos\left\{\frac{2\pi y'}{\lambda} + \phi\right\} - \cos(\phi)\exp\left\{-\left(\frac{\pi\sigma}{\lambda}\right)^2\right\}\right]$$

Where $\sigma = 4$ pixels was a constant that defined the standard deviation of the Gaussian envelope (which determines the spatial extent of the RF); $\gamma = \frac{1}{\sqrt{2}}$ was a constant that defined the aspect ratio of Gaussian envelope (which determines the ellipticity of the RF); $\lambda = 6$ pixels was a constant that defined the wavelength of the sinusoid; ϕ was the phase of the sinusoid; and $x' = x\cos(\theta) + y\sin(\theta)$ and $y' = -x\sin(\theta) + y\cos(\theta)$ where θ defined the orientation of the RF. A family of 32 Gabor functions with eight orientations ($\theta = 0$ to 157.5^o in steps of 22.5^o) and four phases ($\phi = 0^o$, 90^o, 180^o, and 270^o) were used to define the RFs of the neurons in the model. The weights were separated into distinct ON and OFF channels which represented the positive and negative parts of the Gabor function using separate sets of non-negative weights. The cross-correlation and convolution performed in Equations 2 and 3 mean that neurons with these RFs are reproduced at every pixel location in the image.

The size of the RF of a model neuron is specified, above, in pixels. This value should have a direct linear relationship with the size of the RF of a cortical cell measured in degrees of visual angle (ignoring changes in V1 RF size with eccentricity, which is not considered in the model). Similarly, the size, in pixels, of a visual stimulus presented to the model should have the same direct linear relationship with the size of a visual stimulus used in psychophysical experiments. To enable a more direct comparison between the simulation results presented below and the psychophysical data, the parameters of the stimuli used in the simulations are reported in degrees, where a scale factor of 0.16 degrees of visual angle being equivalent to 1 pixel has been used. This scale factor is purely arbitrary, and has been chosen to fit the simulation data to the psychophysical results. For clarity, the symbol ^{o} will be used to refer to the tilt of orientation defined stimuli (measured clockwise from the vertical), while the word "degrees" will be used to refer to the size of stimuli in terms of visual angle.

The results produced by the model are sensitive to the spacing between, and the length of, the texture elements used to define the stimuli. Similarly, results for human observers are also sensitive to these parameters, see sections 3.2 and 3.6. Hence, certain results for the model are dependent on the spacing between the centers of the texture elements and the dimensions of the texture elements used to create the stimuli. Except for those experiments where the spacing or length of the elements is being varied (sections 3.2 and 3.6) all other results have been generated using only two sets of stimulus parameters: (1) a bar length of 1.2 degrees, width of 0.12 degrees, and a spacing of 1.6 degrees (corresponding to the stimulus parameters used in Zhaoping & May, 2007); (2) a bar length of 1.5 degrees, width of 0.15 degrees, and a spacing of 2.5 degrees.

2.5 Calculation of the Saliency Index

At the start of each simulation the prediction neuron responses ($\mathbf{Y}$) were initialized to zero. Then equations 2, 3 and 4 were applied recursively for 10 iterations (results were not sensitive to the exact number of iterations used). At the end of this period the values of $\mathbf{E}$ were recorded at all image locations except those within 20 pixels (3.2 degrees) of the edges of the image (these values were ignored to avoid any possible edge effects). Since each image location is represented by two error neurons (one from the ON-channel and one from the OFF-channel) the maximum of these two values was recorded and used to determine the saliency at that location.

The absolute value of $\mathbf{E}$ is uninformative since an image region with a high error will not necessarily be salient if other regions of the image have similar or higher error. Hence, it is the relative strength of the error signal within a "target" region of the image compared to the strength of the error signal across the rest of the image that is important in determining the saliency of the target region (Zhaoping & May, 2007). As a measure of this relative saliency the recorded error values were used to calculate a saliency index, SI, as follows:

$$SI = f\left\{\frac{\hat{E}_{target} - \hat{E}_{perif}}{\hat{E}_{target} + \hat{E}_{perif}}\right\}$$

Where $\hat{E}_{target}$ was the maximum response across all error-detecting neurons within the target region, $\hat{E}_{perif}$ was the maximum response across all error-detecting neurons across the rest of the image (excluding the edge regions), and $f\{v\} = sign(v)\sqrt{\|v\|}$ provided a nonlinear enhancement of small differences in error. The value of SI can vary from -1 to $+1$ with values less than (or greater than) zero indicating that the error in the target region was less than (more than) the error away from the target region, and hence, that the saliency of the target region was less than (more than) the saliency of the surrounding region.

The regions within which error-detecting neuron responses contributed to the calculation of $\hat{E}_{target}$ and $\hat{E}_{perif}$ for a typical texture segmentation stimulus, a typical feature search stimulus, and a typical contour integration stimulus are illustrated in Figure 2. For a texture segmentation stimulus (Fig. 2a) the saliency of the border was presumed to be related to the relative strength of the error-detecting neuron responses near and far from the border. A feature search stimulus can be considered as an extreme case of a texture segmentation stimulus, in which one texture is made up of a single element (Zhaoping & May, 2007). Hence, for a feature search stimulus (Fig. 2b) the saliency of the unique feature was presumed to be related to the relative strength of the error-detecting neuron responses near and far from the target/distractor border. Similarly, a contour could be considered to be a one element wide texture. Hence, for a contour stimulus (Fig. 2c) the saliency was calculated by comparing the strength of the error-detecting neuron responses near and far from the border formed by the contour. In each case the size of the target region was defined to be d pixels on either side of the texture border (for texture segmentation tasks), or within d pixels of the center of the target stimulus (for feature search tasks), or within d pixels of the elements forming the contour (for contour integration tasks). The value of d was taken to be large enough to encompass one texture element either side of the border, but had a maximum value of 21 pixels (3.36 degrees), to prevent it becoming unreasonably large in widely spaced textures. Results were not sensitive to the exact value used. For experiments where the stimuli were generated randomly, the saliency index is the mean calculated from

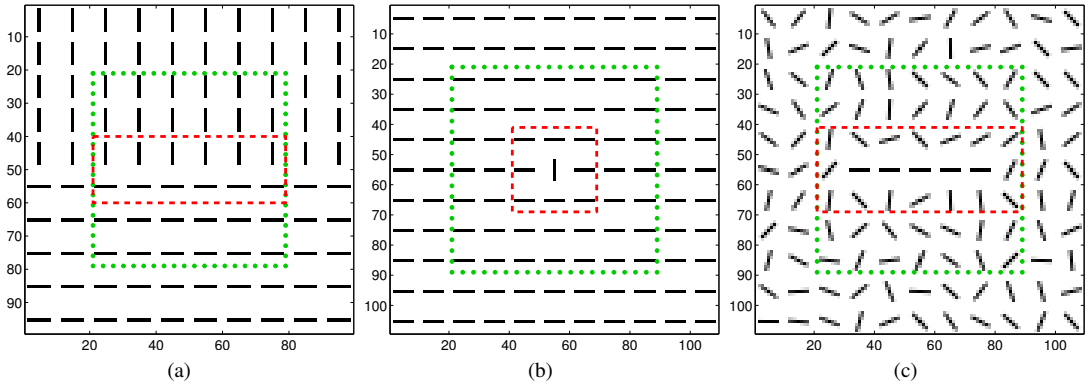

Figure 2: Calculation of the saliency index for (a) a texture segmentation stimulus, (b) a feature search stimulus, and (c) a contour integration stimulus. Each figure shows a pattern of black oriented bars on a white background. The dashed and dotted lines do not form part of the stimulus, but indicate the borders between different regions of the stimulus used to calculate the saliency index. The values of the error-detecting neurons at spatial locations within the dashed box in each image were used to calculate $\hat{E}_{target}$, while the values of the error-detecting neurons within the dotted box, but outside the dashed box, were used to calculate $\hat{E}_{perif}$. The error-detecting neurons outside the dotted box were ignored to avoid any possible edge effects.

20 trials with different randomly generated stimuli. Except where explicitly stated otherwise, the icons used to represent the stimulus configurations in each figure show only the central, 51 by 51 pixel (8.16 by 8.16 degree), portion of the actual images presented to the model.

Note that the calculation of SI described above, is a post-hoc analysis of the output of the model that provides a convenient numerical summary of the result. Importantly, the PC/BC model knows nothing of the location of the target, or the different regions of the image that contribute to $\hat{E}_{target}$ and $\hat{E}_{perif}$.

In psychophysical experiments a number of different measures of saliency have been used: reaction times for the detection of a target; accuracy in the detection of a target; the slope of the function relating reaction times to the number of distractors in a visual search array; the preferred targets for eye movements; the saliency of a target defined by one parameter (e.g., orientation) relative to another target defined by a different parameter (e.g., luminance); and the detectability threshold measured by adjusting the parameter under study (such as orientation). The saliency index calculated for the model should be related to all these psychophysical measures. For example, SI should be inversely related to reaction time, and directly related to accuracy of response. However, the relationship between SI and psychophysical measures of saliency need not be linear. For example, there is a minimum time required for a subject to react, and hence, once the SI of the target has reached a certain value increasing the SI further is unlikely to lead to further improvement in reaction times. Furthermore, a target with low SI might still be found reasonably efficiently. For example, if the target region contained the second highest error in the image, the SI of the target region would be negative. If we presume that the strength of the error at each location influences the probability with which a serial search mechanism visits that location, then the target location (having the second highest error) might still be located rapidly. In another condition, there might be many non-target locations giving rise to an error higher than the target region. In such a case, a serial search method might take a long time to visit all these higher error locations before examining the target. However, in both the above cases the saliency index might be the same. Hence, a zero or negative SI suggests less efficient search than a positive SI, but this less efficient search could have a very large range of efficiencies.

2.6 Code

Software, written in MATLAB, that implements the PC/BC model described above and performs the simulation experiments described below, is available at http://www.corinet.org/mike/Code/v1_saliency.zip.

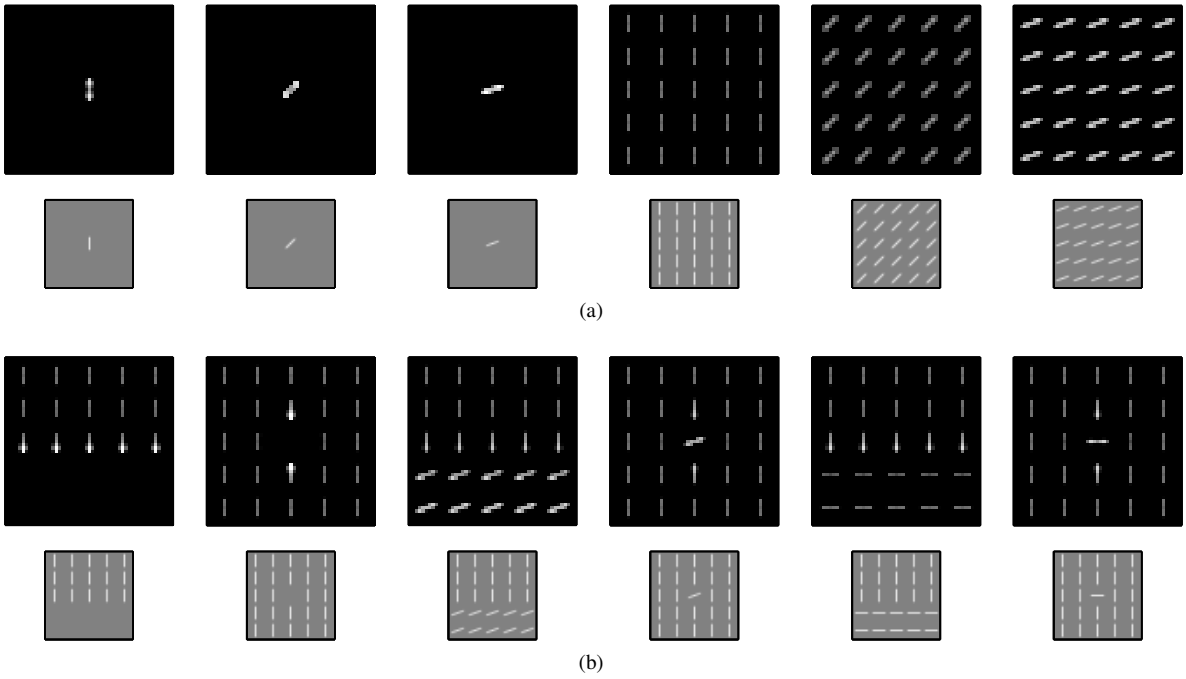

Figure 3: (a) The response of the error-detecting neurons in the PC/BC model to isolated bars and uniform textures formed from oriented bars. The lower row shows the central part of the image presented to the PC/BC model. The first three images are of isolated bar stimuli with orientations 0^o, 45^o, and 70^o. The last three images are uniform textures composed of identical bar stimuli oriented at 0^o, 45^o, and 70^o. The top row shows the strength of the response of the error-detecting neurons (the maximum of $\mathbf{E}_{ON}$ and $\mathbf{E}_{OFF}$ at each pixel). Black pixels show responses of $\leq \frac{2}{\psi}$, while lighter pixels show larger responses. (b) The response of the error-detecting neurons in the PC/BC model to non-uniform textures. The format of the figure, and the gray-scale used to represent the prediction errors, is identical to that used in (a).

3 Simulation Results Without Cortical Feedback

This section describes results obtained using the PC/BC model of V1 when there is no feedback from extrastriate areas (*i.e.*, when all elements of $\mathbf{A}$ are equal to zero and equation 4 can be ignored; see section 2.3). The model is applied to simulating saliency effects in simple luminance defined patterns like those typically used for psychophysical experiments on perceptual saliency. Simulations of texture segmentation and feature search are presented.

The model provides the same explanation for all the results presented below. It proposes that saliency is high in regions of the image which are poorly represented using an over-complete set of Gabor basis functions. To help understand why certain image regions are poorly represented, and hence salient, consider a single active prediction neuron. This neuron will contribute to reducing the error-detecting neuron responses at all locations from which it receives non-zero afferent weights. If the response of the prediction neuron is being driven by a particular feature of the input stimulus, it will reduce the error associated with that feature. However, if the RF of the prediction neuron is larger than that stimulus, it will also reduce the error at neighboring locations. Hence, the error at image location a will be effected by the presence of stimuli at neighboring locations that can drive the responses of prediction neurons with RFs that overlap a.

Isolated image features are therefore generally associated with high prediction error, and hence, high saliency. For example, an isolated bar stimulus is not particularly accurately represented using Gabor functions, and hence, the responses of the error-detecting neurons at image locations near the bar are high. This is true for all bar orientations, as is illustrated by the first three columns of Fig. 3a. However, when a number of oriented bar elements are combined together to form a texture, this can reduce the error generated by, and hence, the salience associated with elements in the texture (as is illustrated by the last three columns of Fig. 3a). This occurs when the spacing between co-aligned texture elements is smaller that the RF size of the prediction neuron RFs. If that is the case, then prediction neurons representing one element of the texture will generate a reconstruction of the stimulus that is non-zero at locations occupied by neighboring texture elements. This additional contribution to

the reconstruction from neighboring prediction neurons has the effect of reducing the error.

The reduction in prediction error is particularly strong when neighboring texture elements are collinear. This is due to oriented bar stimuli activating prediction neurons with the corresponding orientation preference, and the fact that these prediction neurons have RFs that are elongated along that orientation. Hence, as illustrated in the fourth column of Fig. 3a, when texture elements are placed at locations defined by a square grid, error is relatively small for a texture composed of elements at 0^o. The same is true for a texture composed of elements at 90^o, result not shown. This is because the spacing between collinear elements is smallest in these textures. The reduction in error is weaker, but still considerable, for a texture composed of elements at 45^o (see Fig. 3a, column 5). The same is true for a texture composed of elements at -45^o, result not shown. However, for texture elements placed at locations defined by a square grid, the reduction in error is weak in textures composed of elements at other angles, as the spacing between collinear elements is large compared to the size of each prediction neuron's RF (see the right-hand column of Fig. 3a).

If a texture is not regular (*e.g.*, if there are elements missing or at a different orientation to their neighbors), then there will be locations where some texture elements have no collinear neighbors: at the edges of the texture, or where there is a bar with a unique orientation. At these locations the error will be high (see Fig. 3b). In the following sections, it is shown that these effects, caused by the presence or absence of neighboring texture elements, result in prediction errors that are consistent with psychophysical measures of the perceptual salience across a wide variety of different stimuli.

3.1 Texture Segmentation from Orientation Contrast

For textures composed of oriented bars, segmentation generally becomes more efficient as the orientation contrast between the elements making up two textured regions is increased (Nothdurft, 1985, 1991, 1992; Wolfson & Landy, 1995). The model is mostly consistent with this data, with the saliency of the border generally increasing with orientation contrast (Fig. 4a). However, it has been noted that orientation contrast is not the whole story (Ben-Shahar & Zucker, 2004b; Wolfson & Landy, 1995) and in particular, that human segmentation performance improves when texture elements align with the border between the two textured regions (Ben-Shahar & Zucker, 2004b; Giora & Casco, 2007; Nothdurft, 1992; Olson & Attneave, 1970; Wolfson & Landy, 1995). Hence, for a 90^o orientation contrast, segmentation is more efficient when the texture elements are oriented at 0^o and 90^o to the border rather than at $+45^o$ and -45^o to the border (Giora & Casco, 2007; Wolfson & Landy, 1995). The results of the model are consistent with this observation, as is shown in Fig. 4a by the higher saliency at an orientation contrast of 90^o for the results plotted with the diamond and triangular markers compared with those plotted with circular and square markers. The influence of this configural effect is so strong that the efficiency of human texture segmentation is greater for textures oriented at 0^o and 45^o to the border, and (to a lesser extent) for textures oriented at 90^o and 45^o to the border, than for textures oriented at $+45^o$ and -45^o to the border (Wolfson & Landy, 1995). The results of the model are also consistent with this data, as is shown by the higher saliency of the results plotted with the diamond and triangular markers at a 45^o orientation contrast compared with the results plotted with circular and square markers at a 90^o orientation contrast.

The fall in saliency at an angle of 67.5^o seen for the results plotted with the triangular markers is due to a regular texture with elements at this angle being poorly represented (this is the case for many textures with oblique orientations, see the right-hand column of Fig 3a for an example). Hence, in this condition there are high reconstruction errors for all elements of this texture, both within and outside the border area, which reduces the relative saliency of the border.

Texture segmentation generally becomes less efficient as the orientation contrast between the elements within each region is increased (Nothdurft, 1992; Wolfson & Landy, 1995). Fig. 4b shows simulation results equivalent to those shown in Fig. 4a, when the orientation within each texture varies by up to $\pm 15^o$ from the nominal orientation. The heterogeneity of the textures causes a reduction in the saliency of the border across all conditions. This results from there being elements outside the border region which have neighbors that are only roughly collinear. These elements are thus poorly represented by the PC/BC model, reducing the relative saliency of the border. It is only when the orientation contrast at the border is high, that elements that are most poorly represented by the PC/BC model are always located next the border. Note that the saliency index at 0^o orientation contrast is less than zero. This is to be expected, as the location with the highest error will be randomly determined by the orientations chosen for neighboring texture elements. Since there are more texture elements outside the border area, there is a higher probability that the highest error will be away from the border, and hence, that the saliency of the border region is negative.

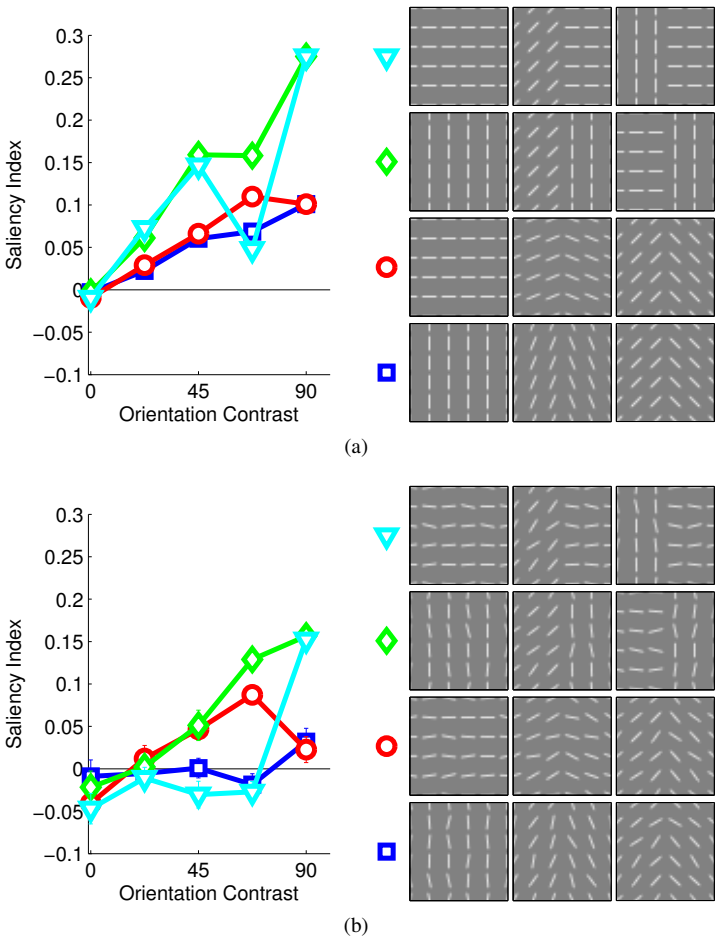

Figure 4: Saliency of the border between orientation defined textures as a function of orientation contrast. (a) The images presented to the model were made up of two uniform textures divided by a vertical border. Each texture was created from oriented bars. The saliency index of the border region is shown as a function of the difference in orientation (orientation contrast) of the texture elements making up the two textures. Results are shown for four configurations of textures, as illustrated by the icons on the right-hand side of the figure. From left to right the icons show stimuli for 0^o, 45^o and 90^o orientation contrast in each of the four configurations. Square and circular markers show results when the elements in each texture made the same angle with the border, starting with both sets of texture elements being parallel to the border in the case of the square markers, and starting with both sets of texture elements being perpendicular to the border in the case of the circular markers. Diamond markers show results when the elements in one texture were always parallel to the border. Triangular markers show results when the elements in one texture were always perpendicular to the border. (b) Saliency of the border between orientation defined textures as a function of orientation contrast in the presence of within texture heterogeneity generated by randomly choosing the orientation of each texture element from a uniform distribution of angles in the range $+15^o$ to -15^o around the nominal orientation of the texture. Error bars show the standard error of the mean over the twenty trials performed for each condition. Note that many error bars are smaller than the corresponding markers, and are hence, not visible. The format of the diagram is otherwise identical to (a) as described above.

3.2 Texture Segmentation with Varying Element Spacing

The ability of human observers to distinguish two textures is known to vary with both the spacing of the texture elements and the length of the texture elements (Nothdurft, 1985). Specifically, for a fixed length of bar the detectability of the border was found to be greatest at small element spacings and fell for larger spacings (Nothdurft, 1985). The reduction in detectability happened at larger spacings for bars of longer length (Nothdurft, 1985). Similar results for the model are shown in Fig. 5a. For a fixed spacing, the ability of human observers to detect the texture border increased with increasing length of the elements making up the textures (Nothdurft, 1985). The in-

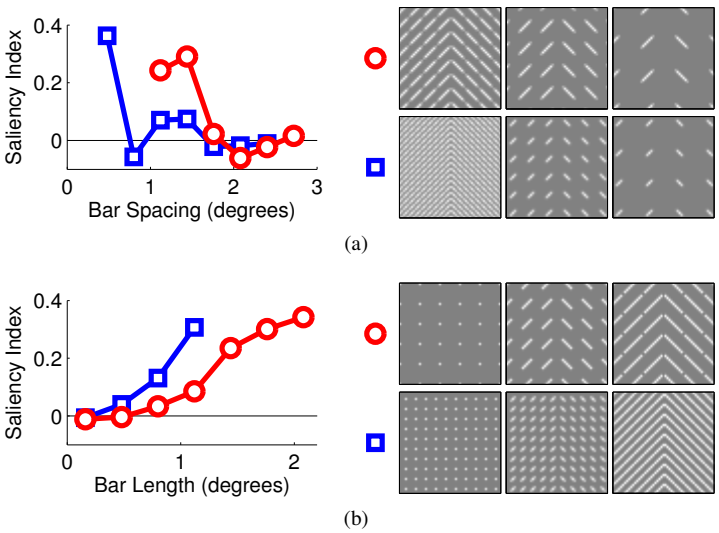

Figure 5: Saliency of the border between orientation defined textures as a function of spacing between, and length of, the texture elements. (a) The effect of changing the spacing between the centers of the texture elements for elements 0.65 degrees long (square markers) and elements 1.3 degrees long (circular markers). (b) the effect of changing the length of the texture elements for textures with a 1.17 degree spacing between the centers of the elements (square markers) and for textures with a 2.33 degree spacing between the centers of the elements (circular markers).

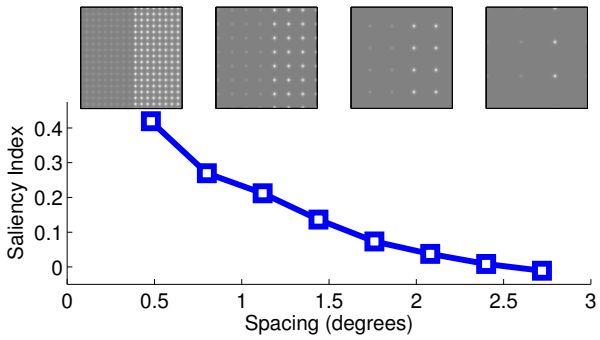

Figure 6: Saliency of a luminance defined border between two textures as a function of texture element spacing. Both textures are made up of square elements (of width 0.22 degrees), but the left-hand texture is dimmer than the right-hand texture. The icons show, from left to right, spacings of 0.48, 1.12, 1.76, and 2.72 degrees.

crease in detectability occurred at shorter bar lengths for more closely packed texture elements (Nothdurft, 1985). Similar results for the model are shown in Fig. 5b. Both sets of simulation results are explained by the increase in distance between collinear texture elements caused by increasing the inter-element spacing or by decreasing the element length. When the distance between collinear elements increases, the prediction error increases. Hence, in these condition there are high reconstruction errors for all elements of both textures, both within and outside the border area, which reduces the relative saliency of the border. The non-monotonic reduction in saliency with increased spacing between the texture elements seen in Fig. 5a is due to interactions between the reconstructions generated by neighboring prediction neurons with *parallel* orientation preferences. Whether the contribution from such neighboring prediction neurons helps reduce the reconstruction error or not will depend on whether of not the wavelength of the Gabor function matches the spacing of the texture elements.

The effect of spacing is not unique to orientation defined textures. Nothdurft (1985) also showed that for luminance defined textures, the detectability of the border decreased (a greater difference in luminance was required to make the border visible) as the spacing of the elements increased (Nothdurft, 1985). In the model, the saliency of the border region between two textures that differ only in luminance also decreases with increasing spacing, Fig. 6. In the PC/BC model, the strength of the response of a prediction neuron is proportional to the strength

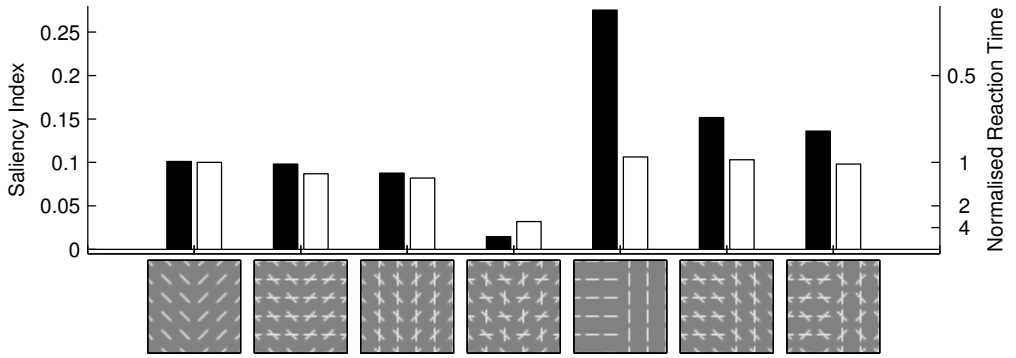

Figure 7: Saliency of the border between orientation defined textures in the presence and absence of a superimposed task-irrelevant texture. Unfilled bars show the reciprocal of the reaction times measured in the psychophysical experiments (averaged over four subjects) and normalized by the reaction time for the border defined by texture elements at $\pm 45^o$, *i.e.*, the reaction time to the stimulus used in the first column (data taken from Fig. 5 in Zhaoping & May, 2007). The scale is shown on the right-hand y-axis. Filled bars show the saliency index of the border region measured in the model. The scale is shown on the left-hand y-axis. The first column shows the saliency of a border defined by texture elements at $\pm 45^o$, and the next three columns show the effects of superimposing on this stimulus regular, task-irrelevant, textures. The fifth column shows the saliency of a border defined by texture elements at 0^o and 90^o, and the next two columns show the effects of superimposing on this stimulus regular, task-irrelevant, textures. Note that in contrast to the experiments described in section 3.1, Zhaoping & May (2007) did not find that detection of a border was easier when the texture elements were oriented at 0^o and 90^o to the border (column 5) rather than at $+45^o$ and -45^o to the border (column 1).

of the input it receives. Hence, strong inputs will generate strong responses, and weak inputs will generate weak prediction neuron responses. If a prediction neuron receives both strong and weak inputs from different parts of its RF, it will produce an intermediate strength of response, which will necessarily underestimate the strength of the strong inputs in the reconstruction, leading to high errors. Hence, in these simulations, at small spacings, the highest errors occur along the border between the regions with the different luminances. As spacing increases, this effect diminishes due to the elements with different luminances becoming increasingly represented by distinct prediction neurons.

3.3 Texture Segmentation with Superimposed Irrelevant Textures

In order to test predictions of the V1 saliency map hypothesis, Zhaoping & May (2007) performed a series of psychophysical experiments to assess the effects of task-irrelevant features on visual search and texture segmentation. Simulations of their results from the visual search experiments are presented in Section 3.7, while simulations of their results from the texture segmentation experiments are presented in this section. Fig. 7 shows the stimuli used in these experiments, and the corresponding saliency index measured from the model, filled bars, together with the reciprocal of the reaction times measured by Zhaoping & May (2007), unfilled bars. It can be seen that for a border defined by texture elements at $\pm 45^o$ the superposition of a regular horizontal or vertical texture slightly reduces the efficiency of border detection in the human subjects, and the saliency measured in the model (compare columns 2 and 3 with column 1 of Fig. 7). In the model, this is due to competition between prediction neurons representing superimposed texture elements. This slightly reduces the responses of those prediction neurons in comparison to the case when only one bar is presented at each location. This reduction in prediction neuron response strength, increases the reconstruction error. This small increase in reconstruction error, both within and outside the border area, slightly reduces the relative saliency of the border.

The superposition of a texture in which the orientation of the elements alternates between horizontal and vertical has a much larger effect, markedly reducing the detectability of the border for the human observers and reducing the saliency index measured from the model (compare column 4 with column 1 of Fig. 7). In the model, this is due to the irrelevant texture elements not having collinear neighbors. The reconstruction error for these elements is thus greatly increased, which further reduces the relative saliency of the border.

For a border defined by texture elements at 0^o and at 90^o the efficiency with which human subjects could detect the border was relatively unaffected by the superposition of either a uniform texture (at -45^o) or a texture with alternating elements (at $\pm 45^o$). In the model, the saliency index is reduced by these task-irrelevant textures,

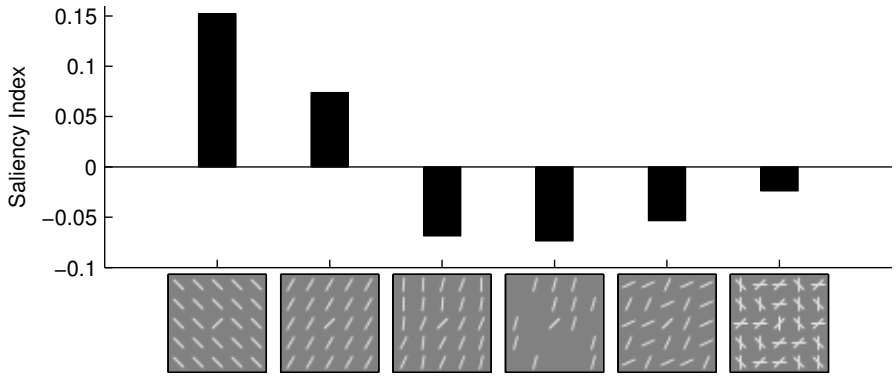

Figure 8: Saliency of the border around an orientation defined singleton in a search array. Results are shown for reducing the orientation contrast between the target and the distractors (columns 1 and 2), introducing heterogeneity in the distractors (columns 3 to 5) and for a target defined by the conjunction of two orientations (right-hand column).

but remains high in each case (compare columns 6 and 7 with column 5 of Fig. 7). Saliency was reduced in the model for the same reason (given above) that it was reduced in the results shown in columns 2 and 3 of Fig. 7 (note that for a uniform texture of -45^o, and for an alternating texture with elements at $\pm 45^o$, elements have the same number of collinear neighbors). The greatest discrepancy between the simulation results and the psychophysical results is for the condition shown in column 5 of Fig. 7. In other psychophysical experiments (as described in section 3.1), the saliency of the border between texture elements oriented at 0^o and 90^o was found to be greater than the saliency of the border between texture elements oriented at $+45^o$ and -45^o. This effect was not found by Zhaoping & May (2007).

3.4 Visual Search Efficiency

Human performance in visual search tasks shows a range of efficiencies, ranging from fast and accurate ("parallel" or "pop-out") searches at one extreme, to slow and inaccurate ("serial") searches at the other (Wolfe, 1998). Generally, search becomes less efficient as the similarity between the target and distractors increases, or the similarity between the distractors decreases (Duncan & Humphreys, 1989; Wolfe, 1998; Wolfe & Horowitz, 2004). Hence, for orientation defined stimuli, the search for a target which has a large orientation difference to the distractors is efficient, but the efficiency of the search is reduced as the orientation contrast between the target and the distractors decreases (Wolfe, 1998; Wolfe & Horowitz, 2004). In the model, the saliency index for the region around a uniquely oriented target (at 45^o) falls as the orientation contrast of the distractors is reduced from 90^o to 15^o (Fig. 8, first and second columns). Feature search becomes much less efficient when the distractors are heterogeneous (Duncan & Humphreys, 1989; Wolfe, 1998, 2003; Wolfe et al., 1992; Wolfe & Horowitz, 2004). In the model, the target becomes less salient when the distractors are made heterogeneous either by allowing a range of orientation contrasts (between 45^o and 15^o), by decreasing the uniformity of the background texture (by randomly selecting elements at an orientation contrast of 30^o to be present with probability 0.5), or by randomly choosing the distractors to be tilted either at $+20^o$ or at -20^o from the orientation of the target (Fig. 8, third to fifth columns). In all these simulations, the target consists of a bar with a unique orientation. Like the examples shown in Fig. 3, this bar is poorly represented by the PC/BC model of V1, and is hence salient when the surrounding texture is uniform. However, all the manipulations that increase the heterogeneity of the distractors, result in the distractors containing elements that are poorly represented by the PC/BC model (elements with fewer or more poorly aligned collinear neighbors). The distractors thus contain elements that are associated with similar, or higher, reconstruction errors than the target and the target ceases to be salient.

Search is also inefficient for targets defined as the conjunctions of two orientations (Wolfe, 1998). In the model, the target becomes less salient than the distractors for such a conjunction search (in this example, the target is a conjunction of 0^o and 45^o bars among distractors consisting of 90^o and 45^o bars together with 0^o and -45^o bars, with one of the two distractors selected at random for each location in the background texture, Fig. 8, right most column). The distractors in this condition contain two superimposed, irregular, textures many elements of which do not have collinear neighbors, and are hence salient.

Generally, as the similarity between the target and distractors increases, human search performance becomes less efficient (Duncan & Humphreys, 1989; Wolfe, 1998; Wolfe & Horowitz, 2004). However, a 50^o target among

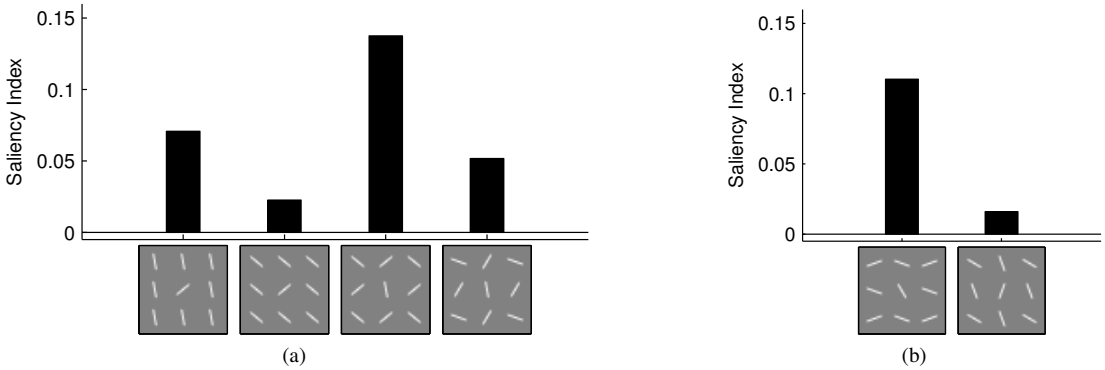

Figure 9: Saliency of the border around an orientation defined singleton in a search array. (a) Results are shown for a 50^o target among -10^o distractors (column 1), a 50^o target among -50^o distractors (column 2), a -10^o target among -50^o and $+50^o$ distractors (column 3), and a -10^o target among -70^o and $+30^o$ distractors (column 4). (b) Results are shown averaged over 16 trials with different combinations of target and distractor orientation (as used in Wolfe & Friedman-Hill, 1992). In each case two orientations of distractors were used. For column 1, the distractors were symmetrical across the vertical axis. For column 2, 50^o was added to the orientation of the target and distractors so that the distractors were symmetrical about an oblique 50^o angle. One example stimulus is shown in each icon: a -30^o target among $+70^o$ and -70^o distractors (column 1), and a $+20^o (= -30^o + 50^o)$ target among $-60^o (\equiv 70^o + 50^o)$ and $-20^o (= -70^o + 50^o)$ distractors (column 2).

-50^o distractors has been found to be harder to find than a 50^o target among -10^o distractors (Wolfe, 1998), despite the increased similarity in orientation between the target and distractors in the latter case. This effect can be simulated by the model, as shown in Fig. 9a, columns 1 and 2. In both cases the target element has no collinear neighbors, and hence, is salient in the model. When the distractors are oriented at -10^o, each distractor has roughly collinear neighbors in close proximity. In contrast, when the distractors are oriented at -50^o the roughly collinear neighbors are more distant. Hence, in the latter case the error in reconstructing the distractors is greater, and the relative saliency of the target is less.

In addition, it has been found that a -10^o target among distractors oriented at both -70^o and $+30^o$ is harder to find than a -10^o target among distractors at -50^o and $+50^o$ (Wolfe, 1998; Wolfe et al., 1992). This is despite the fact that the orientation of the target differs by 40^o from that of the most similar distractor in both cases. The model also succeeds in simulating these effects (Fig. 9a, columns 3 and 4). The explanation provided by the model is that the distractors in the $-70^o/+30^o$ condition all have roughly collinear neighbors, whereas in the $-50^o/+50^o$ condition, distractors do not have collinear neighbors. The distractors are more salient in the latter case, reducing the relative saliency of the target.

Wolfe & Friedman-Hill (1992) found that when two orientations of distractor were used, a target was easier to find when the distractors were symmetrical around the vertical axis than when the distractors were symmetrical around an oblique axis at 50^o. Using an identical set of target and distractor orientations as was used in the psychophysical experiments (Wolfe & Friedman-Hill, 1992), the model successfully simulates this result as the saliency index is higher for vertically symmetrical distractors than for obliquely symmetrical distractors, Fig. 9b. The simulation results suggest that this effect is due to placing the elements of the search array on a square grid with vertical columns and horizontal rows. On such a grid, elements that are symmetrical about the vertical (or horizontal axis) will have roughly collinear neighbors. Hence, in the vertical symmetry condition the distractors have a relatively low salience. However, when the distractors are rotated, at fixed positions in the grid, this produces a texture in which elements do not have collinear neighbors. Hence, in the condition where the elements are rotated by 50^o the distractors are more salient, and hence, the relative saliency of the target is reduced.

Changes in search efficiency have also been observed for targets and distractors which are not straight lines. For example, for the task of finding a circle among distractor circles containing a gap, the search efficiency reduces as the size of the gap in the distractors gets smaller (Treisman & Gormican, 1988; Treisman & Souther, 1985). In the model the saliency of the target also reduces as the gap in the distractors reduces; see the first three columns of Fig. 10. For the task of finding a target line among curved distractors, the search efficiency of human observers reduces as the distractors become less curved (Treisman & Gormican, 1988). The same effect is observed in the model; see the last three columns of Fig. 10. In all these simulations, the distractors are more salient than the target: the error generated by the PC/BC model of V1 is greater for circles with gaps than it is for complete circles, and the

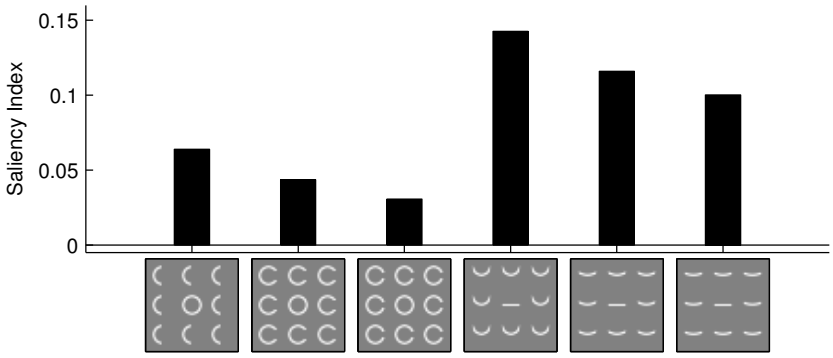

Figure 10: Saliency of the border around a singleton in a search array. Results are shown for a circular target among distractor circles containing a gap (columns 1 to 3). The width of the gap in the distractors decreases from left to right. Results are also shown for a straight bar target among curved distractors (columns 4 to 6). The distractors become less curved from left to right.

error is greater for curved lines than it is for straight ones. However, interactions between the prediction neurons representing the target, and prediction neurons with overlapping RFs that represent the neighboring distractors result in the location with the highest error always being on a distractor that neighbors the target. This makes the boundary between the target and the surrounding texture salient. As the distractors become more similar to the target, so the effect is decreased.

Similar effects on search efficiency have been observed for luminance defined elements. Hence, as the similarity between the target and distractors increases, so search becomes less efficient (Nagy & Sanchez, 1992; Nothdurft, 2006; Pashler et al., 2004). In the model, for a bright target against a dark background, the salience of the target becomes smaller as the luminance of the distractors increases, and hence, becomes more similar to the target (Fig. 11a, first three columns). For bright distractors against a dark background, the salience of a dull target becomes smaller as the luminance of the target increases, and hence, becomes more similar to the distractors (Fig. 11a, columns four to six). Distractor heterogeneity reduces the saliency of the target. Hence, a bright target among medium and dim distractors (Fig. 11a, column 7) is less salient than the same bright target among homogeneously dim or medium distractors (Fig. 11a, columns 2 and 3). Similarly, a dim target among medium and bright distractors has a low saliency index (Fig. 11a, column 8). For colored elements search is found to be particularly inefficient when the target color is chromatically intermediate between the distractor colors (Bauer et al., 1996; D'Zmura, 1991). Similarly, in the model a medium target among bright and dim distractors is less salient than the distractors (Fig. 11a, last column). These results are unaffected by contrast polarity, as shown in Fig. 11b. In the PC/BC model, the strength of the response of a prediction neuron is proportional to the strength of the input it receives. Hence, strong inputs will generate strong responses, and weak inputs will generate weak prediction neuron responses. If a prediction neuron receives both strong and weak inputs from different parts of its RF, it will produce an intermediate strength of response, which will necessarily underestimate the strength of the strong inputs in the reconstruction, leading to high errors. Hence, in these simulations, locations where there are the largest differences in luminance levels within a neighborhood (which includes pixels both on the squares and on the background between the squares) will have the highest errors, and highest salience. Because the V1 model receives input from both the ON and OFF channels of the LGN model, both high and low luminance inputs are treated identically, and the results are unaffected by contrast polarity.

3.5 Asymmetry in Feature Search

Search asymmetries are said to occur when searching for feature a among distractors of type b is more efficient than searching for feature b among distractors of type a (Treisman & Gormican, 1988; Treisman & Souther, 1985; Wolfe, 2001). Simulation results for the model when applied to a number of these search tasks are shown in Fig. 12. Each pair of columns shows the saliency index of the target region when tested with a pair of stimuli in which the roles of target and distractors are reversed. In each case the saliency calculated by the model is in agreement with the efficiency of the corresponding search task for human observers. Specifically:

- A vertical target among tilted distractors is harder to find than a tilted target among vertical distractors (Carrasco et al., 1998; Fahle, 1991; Foster & Ward, 1991; Treisman & Gormican, 1988; Treisman & Souther, 1985; Wolfe, 1998; Wolfe et al., 1992).

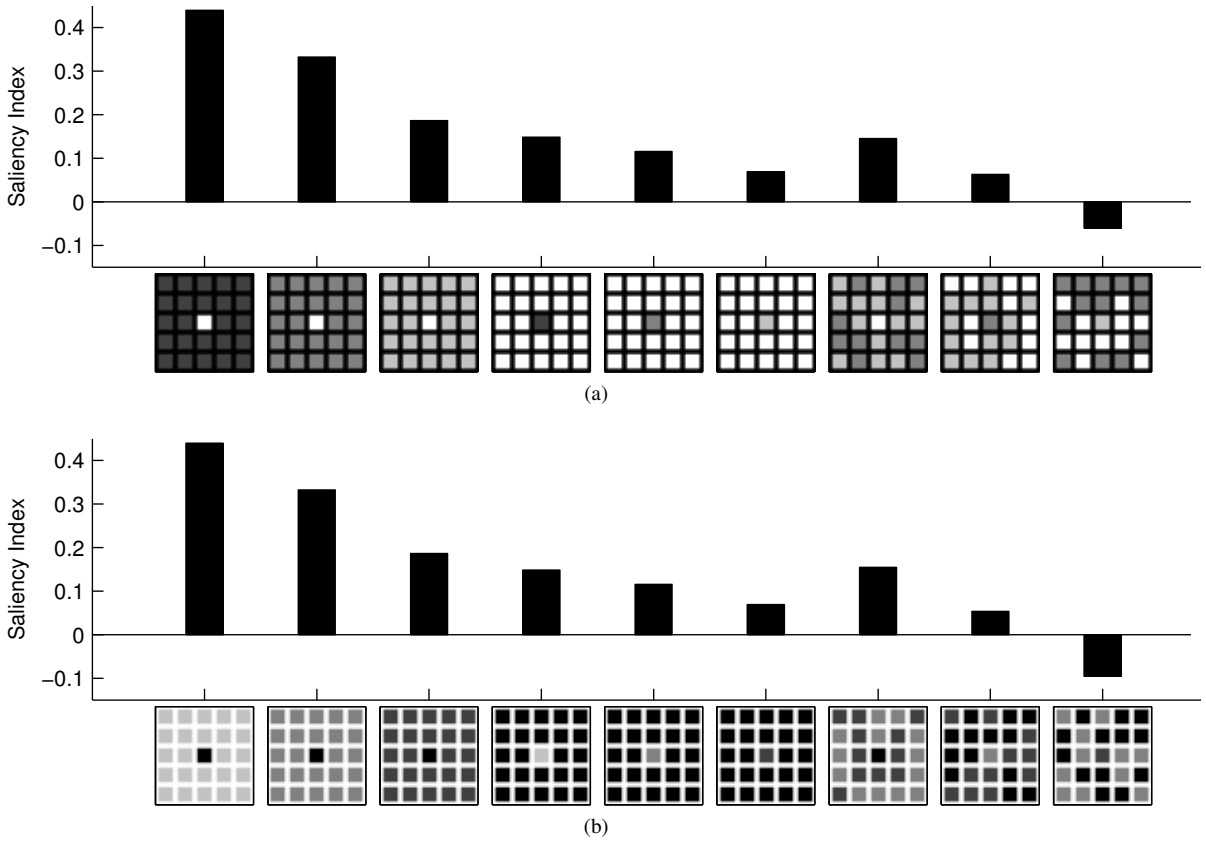

Figure 11: Saliency of the border around a luminance defined singleton in a search array. (a) Results are shown for reducing the contrast between a fixed luminance, bright, target and the distractors (columns 1 to 3), reducing the contrast between a variable luminance target and bright distractors (columns 4 to 6), and for targets with heterogeneous distractors: a bright target among dim and medium distractors (column 7), a dim target among bright and medium distractors (column 8), and a medium target among bright and dim distractors (column 9). (b) Results are shown for stimuli with reversed contrast polarity in comparison to (a).

- A unbroken target bar among distractor bars containing a vernier offset is harder to find than a target bar containing a vernier offset among unbroken distractor bars (Fahle, 1991).
- A target bar among crosses is harder to find than a target cross among bars (Li, 2002), a specific example of the general finding that search for the presence of a feature is more efficient than search for the absence of a feature (Treisman & Gormican, 1988; Treisman & Souther, 1985).
- A pair of converging lines among pairs of parallel lines is easier to find than a pair of parallel lines among pairs of converging lines (Treisman & Gormican, 1988).
- A circle among ellipses is harder to find than an ellipse among circles (Treisman & Gormican, 1988).
- A straight target bar among curved distractor bars is harder to find than a curved target among straight distractor (Fahle, 1991; Treisman & Gormican, 1988; Wolfe, 1998).
- An unbroken target circle among distractor circles containing a gap is harder to find than a broken target circle among unbroken circular distractors (Treisman & Gormican, 1988; Treisman & Souther, 1985).
- A target "O" among distractor "Q"s is harder to find than a target "Q" among distractor "O"s (Saiki et al., 2005; Treisman & Souther, 1985).

For all the simulation shown in Fig. 12a, the higher salience condition occurs when the distractors form a regular texture, in which the elements have nearby, collinear, neighbors. Under these conditions the distractor elements are less salient, increasing the relative saliency of the target. For all the simulation shown in Fig. 12b,

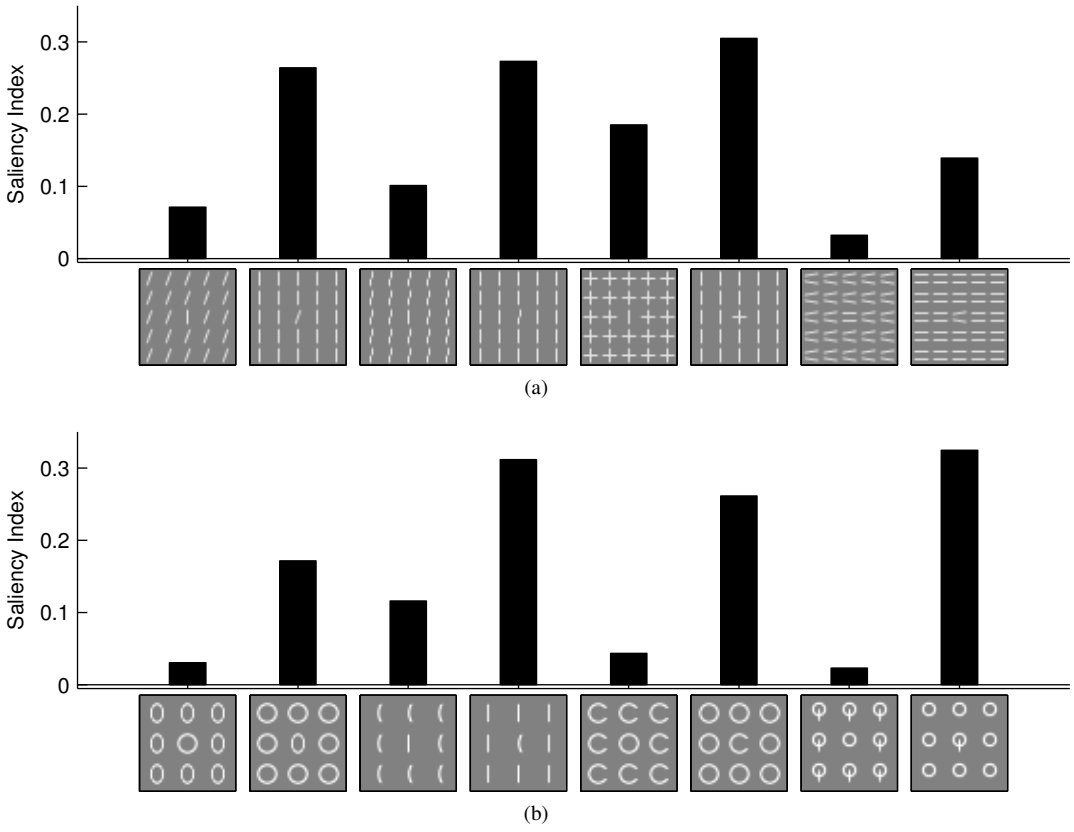

Figure 12: Saliency of the border around a singleton in a search array for pairs of search arrays in which the roles of target and distractor are reversed. (a) a 0^o target among 20^o distractors (column 1), a 20^o target among 0^o distractors (column 2), an unbroken 0^o target among 0^o distractors containing a 1 pixel (0.16 degrees) vernier offset (column 3), a 0^o target containing a 1 pixel (0.16 degrees) vernier offset among unbroken 0^o distractors (column 4), a bar target among cross distractors (column 5), a cross target among bar distractors (column 6), a parallel target among converging distractors (column 7), a converging target among parallel distractors (column 8). (b) a circular target among elliptical distractors (column 1), an elliptical target among circular distractors (column 2), a straight target among curved distractors (column 3), a curved target among straight distractors (column 4), a circular target among distractor circles containing a gap (column 5), a circular target containing a gap among circular distractors (column 6), a circular target among distractor circles with a superimposed bar (column 7), a target circle with a superimposed bar among circular distractors (column 8).

the higher salience condition occurs when the target is a shape that is less accurately represented by the set of Gabor functions used by the PC/BC model of V1. As shown in Fig. 13, reconstruction error is higher for ellipses compared to circles, curved lines compared to straight lines (see Fig. 3a for the straight line stimulus), broken circles compared to unbroken circles, and "Q"s compared to "O"s.

3.6 Feature Search with Varying Element Spacing

Nothdurft (2000) found that for human observers the saliency of an orientation defined target varied as the spacing of the elements in the search array varied. Specifically it was found that the saliency peaked for a spacing of around 1 to 2 degrees of visual angle and decreased both for denser and sparser search arrays (Nothdurft, 2000). Similar results are obtained with the model, Fig. 14. The explanation for these results proposed by the model is identical to that proposed for the equivalent experiments performed on texture segmentation, see section 3.2. Namely, that the individual elements in the texture formed by the distractors become more salient as they become more isolated, due to the lack of overlap with prediction neuron RFs representing collinear elements. Furthermore, as in the texture experiments, the non-monotonic reduction in saliency is a result of interactions, at small spacings, between prediction neuron RFs representing parallel distractor elements.

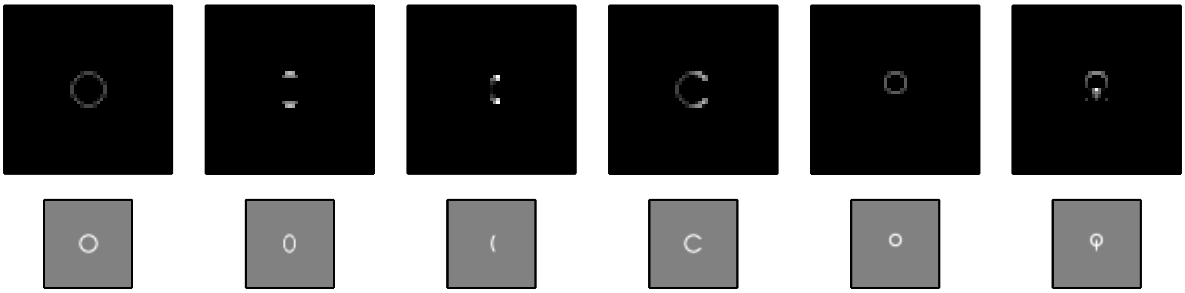

Figure 13: The response of the error-detecting neurons in the PC/BC model to the elements used to create the search arrays shown in Fig. 12b. The lower row shows the central part of the image presented to the PC/BC model. The top row shows the strength of the response of the error-detecting neurons (the maximum of $\mathbf{E}_{ON}$ and $\mathbf{E}_{OFF}$ at each pixel). Black pixels show responses of $\leq \frac{2}{\psi}$, while lighter pixels show larger responses. The gray-scale used to represent the prediction errors is identical to that used in Fig. 3.

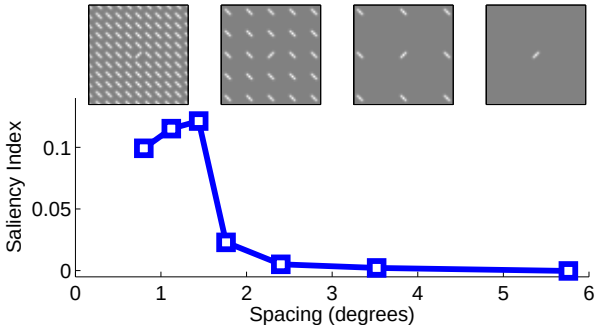

Figure 14: Saliency around an orientation defined singleton in a search array as a function of the spacing between the elements in the search array. The icons show, from left to right, spacings of 0.8, 1.76, 3.52, and 5.76 degrees. The dimensions of the texture elements were 0.66 x 0.15 degrees.

3.7 Feature Search with Superimposed Irrelevant Textures

The effect of superimposing task-irrelevant textures on the search for a unique orientation is shown in Fig. 15. The filled bars show the saliency index for the region around the target measured from the model. The unfilled bars show the reciprocal of the reaction times measured by Zhaoping & May (2007). It can be seen that for a -45^o target among $+45^o$ distractors the superposition of a regular horizontal or vertical texture reduces the efficiency of target detection in the human subjects, and the saliency measured in the model (compare columns 2 and 3 with column 1 in Fig. 15). The superposition of a texture in which the orientation of the elements alternates between horizontal and vertical has an even greater effect in reducing human search efficiency and the modeled saliency index (compare column 4 with column 1 in Fig. 15). For a 90^o target among 0^o distractors, search efficiency for the human subjects is reduced by the superposition of a uniform texture (at -45^o) and by a texture of alternating elements (at $\pm 45^o$). In the model, the saliency index is also reduced by both these the task-irrelevant textures (compare columns 6 and 7 with column 5 in Fig. 15). The model proposes the same explanation for these results as it does for the equivalent experiments performed on texture segmentation, see section 3.3.

3.8 Feature Search with Abrupt Onset

Salience is also affected by the relative timing at which elements appear. For example, search is efficient when the target in a search array appears via an abrupt onset, while the other elements are revealed by removing elements from a mask (Jonides & Yantis, 1988; Yantis & Jonides, 1984). The model simulates these results by showing high salience for an onset target, see column 1 of Fig 16. In a similar experiment, Mounts & Tomaselli (2005) found that reaction times were shorter when an abrupt onset target was flanked by non-onset distractors (*i.e.*, ones that were already present) than when the target was flanked by onset distractors (*i.e.*, that appeared simultaneously with the target). The model also simulates these results, compare columns 2 and 3 of Fig 16. This behavior in the model is due to prediction neuron activity taking time to change in response to changes in the input. An abrupt

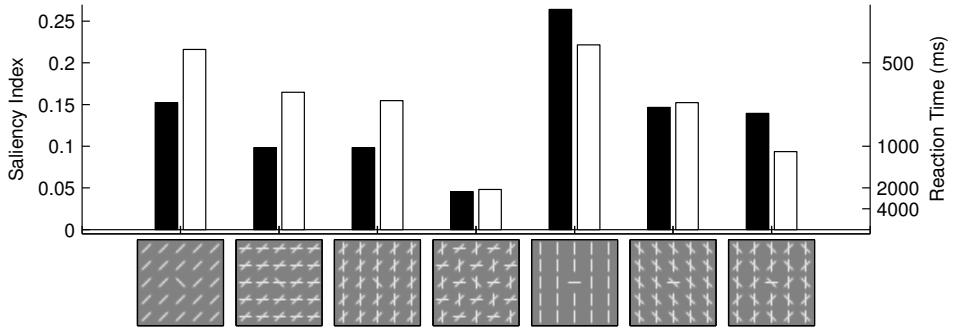

Figure 15: Saliency of a unique target orientation in the presence and absence of a superimposed task-irrelevant texture. Unfilled bars show the reciprocal of the reaction times (averaged over five subjects) measured in the psychophysical experiments (data taken from Table 2 in Zhaoping & May, 2007). The scale is shown on the right-hand y-axis. Filled bars show the saliency index of the border region measured in the model. The scale is shown on the left-hand y-axis. The first column shows the saliency of a -45^o target among distractors at $+45^o$, and the next three columns show the effects of superimposing on this stimulus regular, task-irrelevant, textures. The fourth column shows the saliency of a 90^o target among distractors at 0^o, and the next two columns show the effects of superimposing on this stimulus regular, task-irrelevant, textures.

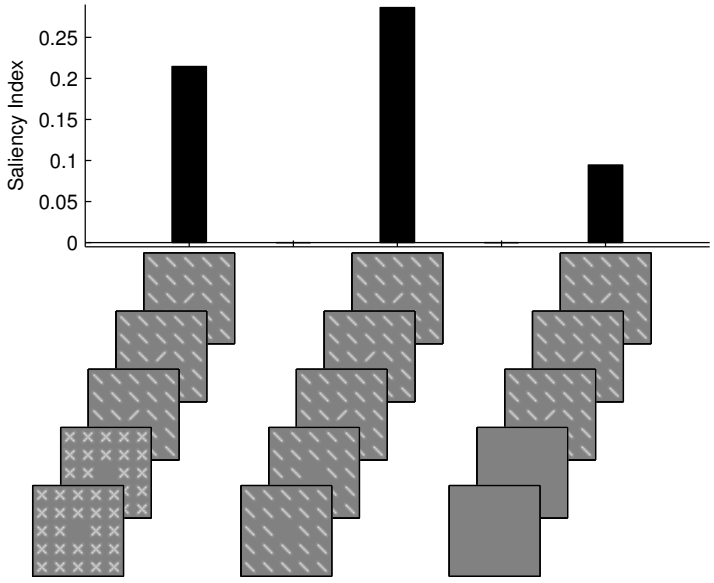

Figure 16: Saliency of the border region around a singleton in a search array as a function of onset of the target and the distractors. The icons show the temporal sequence of images presented to the model (only the images presented at iterations 2, 4, 6, 8 and 10 of the algorithm are shown for clarity). In the first column masks are initially shown at the locations of the distractors, and at iteration six the target abruptly onsets and the distractors are revealed by removing one element from the masks. In the second column the distractors onset at iteration one, and the target onsets at iteration six. In the third column both target and distractors abruptly onset at iteration six.

onset stimulus will thus be particularly salient until the prediction neurons have had time to respond strongly to that stimulus, and hence, provide an accurate reconstruction of it. In contrast, stimuli that have been present for some time will have already generated strong responses in the prediction neurons that represent them, and hence, be represented with smaller reconstruction error. The reduction in saliency for the condition shown in column 1, compared to column 2, of Fig 16 is due to the competition between the prediction neurons representing the mask and distractor stimuli at each location. Such competition results in the prediction neuron responses to the distractors being slightly weaker when the mask is present and for some time after the removal of the mask. Hence, the reconstruction error is slightly higher for the distractors, reducing the relative saliency of the target.

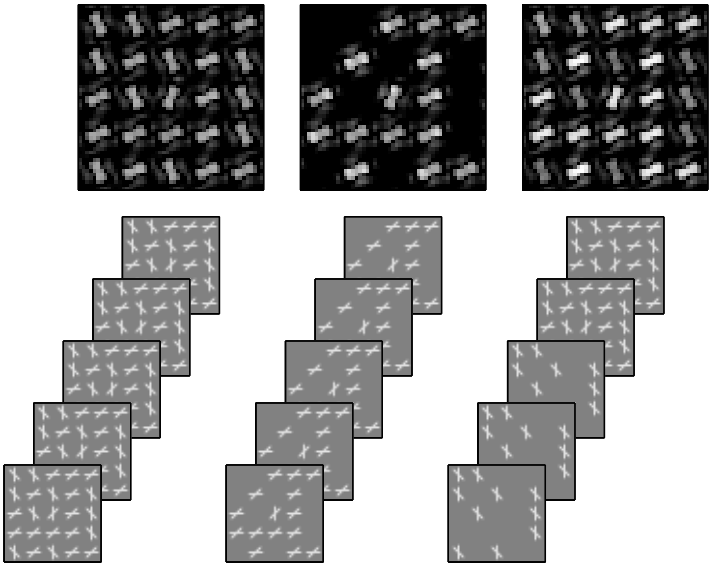

Figure 17: The activation of the error-detecting neurons in response to both old and new elements presented simultaneously (left column), new elements only (middle column), and old and new elements after a preview of the old elements (right column). The strength of the error-detecting neuron responses at each location is indicated by the greyscale: black pixels show an error of $\leq \frac{1}{\psi}$, while lighter pixels indicate locations with higher error. The maximum error across the ON- and OFF-channels is shown. As described in the caption to Fig.16 the icons illustrate the temporal sequence of images presented to the model. Only in the last column does the input change: the new items abruptly onset at iteration eight.

3.9 Preview Search

In a preview search experiment, one set of distractors (the old elements) appears followed, after a time interval, by a second set of elements (the new elements) consisting of additional distractors and the target (Donk, 2005, 2006; Donk & Theeuwes, 2001; Olivers et al., 2006; Theeuwes et al., 1998; Watson & Humphreys, 1997). Search efficiency is similar to that obtained in a standard search task using only the new elements, hence, the old elements have little effect on search efficiency. The model shows similar behaviour, with new elements being more salient than old ones (see Fig. 17). Watson & Humphreys (1997) proposed that the preview benefit could be explained by "visual marking"; a suppression of the initial set of distractors. The model is consistent with this view in the sense that prior exposure to one set of distractors will activate the corresponding prediction neurons. These neurons will generate predictions about the locations/features of those distractors, which will in turn decrease the error generated in response to those distractors. Hence, items added later will be relatively more salient. Donk & Theeuwes (2001) proposed that the preview effect was the result of the abrupt onset of the new elements capturing attention. The model is also consistent with this explanation, since the onset of an element generates high saliency until the prediction neurons have formed a representation of the input. Hence, the PC/BC model is consistent with the idea that both old-element suppression, and new-element attention capture, operate together (Donk, 2005, 2006; Olivers et al., 2006). However, the PC/BC model proposes that both of these behaviors are side-effects of a more general computational principle: predictive coding.

3.10 Change Blindness

If visual stimuli consistently change in predictable ways, it would be possible for the PC/BC model to learn an internal representation of the world that provided predictions of these changes to make such foreseeable changes less salient[2]. However, in the absence of such a model, or in the presence of unpredictable changes to the visual input, the predictive coding model makes the default assumption that the input will remain constant. Hence, sudden changes at specific locations in the input become salient, and lead to the results described above in sections 3.8 and 3.9. If however, the whole image changes rather than only certain parts of it, then all locations in the image will be salient due to the unpredictable changes that have occurred everywhere. This behavior of the model is consistent with change blindness induced in human subjects by a visual disruption or 'flicker' (Rensink et al., 1997, 2000).

[2]Such a model might be encoded by the prediction neurons in an extended PC/BC model of V1, or more likely by prediction neurons modeling higher cortical regions with these higher-level predictions being fed-back to V1.

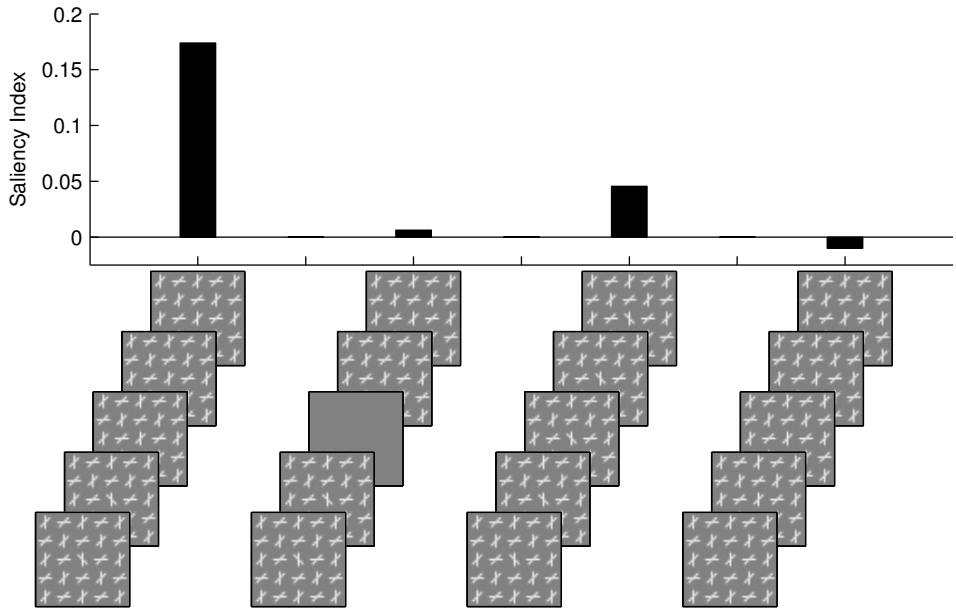

Figure 18: Simulation of change blindness. The icons show the temporal sequence of images presented to the model (only the images presented at iterations 2, 4, 6, 8 and 10 of the algorithm are shown for clarity). In the first column, starting at iteration 6, one bar located at the center of the image changes orientation from -45^o to $+45^o$. In column 2 the same change occurs, however, the changed image is preceded by a blank image shown during iterations 6 and 7. The third and fourth column show results for unchanging inputs. In column 3, the first of the two stimuli used in the preceding experiments is shown throughout the simulation. In column 4, the second of the two stimuli is shown throughout the simulation.

When a stimulus at one location changes, this change is highly salient, as shown in the first column of Fig. 18. However, when the same change in the stimulus is preceded by a briefly presented blank image, the change is not salient, as shown in the second column of Fig. 18. The blank stimulus causes the responses of all prediction neurons to become weak (effectively, the model starts to predict that the input will remain blank). Hence, when the changed image is presented the prediction is poor at all locations occupied by texture elements, and hence, the reconstruction error in the border region around the changed feature is approximately the same as the error at locations outside the border region.

In the simulation described above the first of the two images is the same as that used to generate the result shown in column 4 of Fig. 15. The border region around the singleton in this search array is only weakly salient, as is illustrated in column 3 of Fig. 18 where this image is presented to the model at all iterations. The second of the two images does not contain a singleton, and hence, the border region around the central element is not salient, as is illustrated in column 4 of Fig. 18. Hence, the high saliency caused by the change in input without the intervening blank stimulus (shown in column 1 of Fig. 18) is entirely due to the change in the image, not the saliency of either the image presented before or after the change.

4 Simulation Results With Cortical Feedback

So far all simulations have ignored the influences of cortical feedback connections. In contrast, this section describes results obtained using the PC/BC model of V1 when there is feedback from extrastriate areas (*i.e.*, when elements of $\mathbf{A}$ are not all equal to zero and equation 4 can not be ignored; see section 2.3).

The model provides the same explanation for all the results presented below. Consider two texture segmentation tasks, one of which has a salient border and one of which has a far less salient border (the same stimuli as used in columns 1 and 4 of Fig. 7). The activity of the error-detecting neurons, when there is no top-down excitation, is shown in Fig. 19a. When top-down excitation is provided to the right-hand side of each image, this increases the response of the prediction neurons representing elements on that side of the image. The increased prediction neuron responses lead to a stronger reconstruction of the input and a reduction in the responses of the error-detecting neurons in that part of the image, as shown in Fig. 19b. However, the saliency index of the border

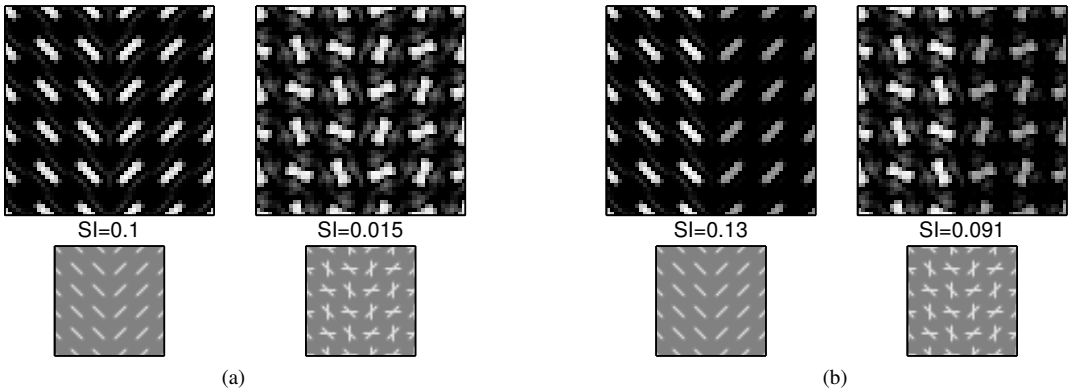

Figure 19: The activation of the error-detecting neurons in response to two texture segmentation stimuli (a) with no cortical feedback, (b) with cortical feedback to the right-hand half of the image. The lower row shows the central part of the image presented to the PC/BC model. The top row shows the strength of the response of the error-detecting neurons (the maximum of $\mathbf{E}_{ON}$ and $\mathbf{E}_{OFF}$ at each pixel). Black pixels show responses of $\leq \frac{1}{\psi}$, while lighter pixels show larger responses. The grey-scale is the same in (a) and (b). In each case the saliency index (SI) of the border region is indicated.

is increased, considerably so in the case of the harder segmentation task. As described in section 3, error, and hence saliency, tends to be high at locations where the regularity of the texture breaks-down. Top-down input emphasizes the change in texture across the border, and hence, increases the saliency.

Several of the simulations described below suggest that feedback from higher-level perceptual representations modulates V1 saliency. One objection to this proposal might be a concern about the time required for information to propagate up and down the cortical hierarchy, and hence, for top-down influences to affect processing (Summerfield & Egner, 2009). However, as with models of cortical hierarchies which only contain feedforward connections (*e.g.*, Riesenhuber & Poggio, 1999; Serre et al., 2007b; Thorpe et al., 2004; Wallis & Rolls, 1997), PC/BC is capable of performing rapid processing via the first feedforward wave of activity traveling up the hierarchy. In situations where the stimulus is sufficiently well encoded by the synaptic weights, this rapid, feedforward only, processing may be sufficient to activate appropriate high-level representations, as appears to be the case for human observers (Keysers et al., 2001; Li et al., 2002; Serre et al., 2007a; VanRullen & Thorpe, 2001). Once these high-level representations are active, this information can just as quickly propagate back down the hierarchy to affect ongoing processing in V1. The prediction errors calculated within higher processing stages might also be output to the thalamus to affect behavior (see Discussion) reducing the time required for higher-level influences on behavior. Furthermore, in situations where there is an expectation about a stimulus before it appears, top-down influences on V1 may be present before stimulus onset, and hence, have an even more rapid influence.

4.1 Contour Detection

When a number of oriented elements align so that they appear to form a contour, they pop-out of a surrounding texture of randomly oriented elements (Field et al., 1993; Hess et al., 2003). The detectability of the contour increases as the path length increases (Li & Gilbert, 2002; Watt et al., 2008). In the model the salience of the border region surrounding a contour also increases with the number of elements forming that contour (see Fig. 20). The model treats contour integration as just another form of texture segmentation, where one texture happens to be one element wide. As shown in Figs. 4a and 4b when one set of texture elements are collinear with the border (*i.e.*, as are the elements forming the contour), the saliency of the border will be maximum when the elements forming the other texture are perpendicular to the border and will be minimum when the elements forming the other texture are parallel to the border. This is consistent with recent psychophysical results showing that when the orientation of the surrounding distractors is parallel to the contour it is much less salient than when the surrounding distractors are perpendicular to the contour (Dakin & Baruch, 2009). When, as is the case in Fig. 20, top-down excitation is provided to the prediction neurons representing the contour the saliency of the locations occupied by these elements decreases, but the error in encoding neighboring elements increases so that the saliency of the border is increased. Such top-down excitation might come from higher-level perceptual representations of contours that have been activated by the presence of multiple co-aligned elements.

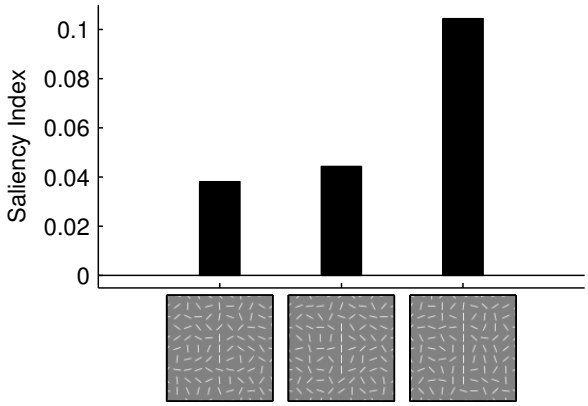

Figure 20: Saliency of the border around a contour embedded in a random texture with cortical feedback targeting the contour. Results are shown for increasing contour length: from left to right contour length equals 3, 5, and 7 elements. In contrast to all other figures, the icons used to represent the stimulus configurations show the central 101 by 101 pixel portion of the actual images presented to the model.

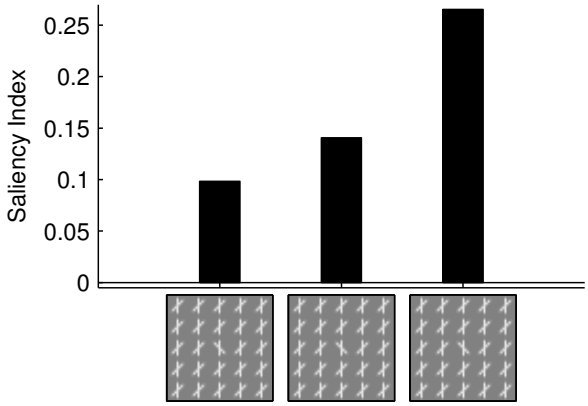

Figure 21: The saliency of the border around a singleton in a search array, with (column 1) no cortical feedback, (column 2) cortical feedback to V1 prediction neurons with orientation preferences of 0^o and -45^o (*i.e.*, the features of the target), (column 3) cortical feedback to V1 prediction neurons with orientation preferences of 0^o and $+45^o$ (*i.e.*, the features of the distractors).

4.2 Feature Search with Novel and Familiar Elements

The reduction in prediction error for familiar shapes would also explain the asymmetry that is observed between efficient searches for novel targets among familiar distractors compared to less efficient searches for familiar targets among novel ones. For example, search is efficient for a mirror reversed letter among the same letters presented normally, but is less efficient for a target letter among mirror-reversed distractors (Frith, 1974; Reicher et al., 1976; Saiki et al., 2005; Shen & Reingold, 2001; Wang et al., 1994; Wolfe, 2001). A simple simulation of this effect is shown in Fig. 21. The stimulus presented to the model contains a unique "shape" made up of bars oriented at 0^o and -45^o among distractors made up of a conjunction of bars oriented at 0^o and $+45^o$. Hence, the target and distractors are mirror-reversed versions of each other. When neither shape is familiar and there is no top-down activation of the V1 prediction neurons, the border region around the target is reasonably salient (Fig. 21, column 1).

When the target shape is familiar, there will be higher-level perceptual representations of this shape which will be activated by its presence in the search array. These higher-level representations will subsequently send top-down activation to V1. Hence, to simulate familiarity for the target, all prediction neurons in the V1 model with an orientation preferences of 0^o or -45^o received top-down activation. The effects of this are shown in Fig. 21, column 2. The "familiarity" of the target increases the saliency of the border region around the target. This is due to increased error at locations occupied by distractors neighboring the target.

When the shape of the distractors is familiar, there will be higher-level perceptual representations of this shape which will be activated by its presence in the search array and which will in turn provide top-down activation to

V1. Hence, to simulate familiarity for the distractors, all prediction neurons in the V1 model with an orientation preferences of 0^o or $+45^o$ received top-down activation. The effects of this are shown in Fig. 21, column 3. The "familiarity" of the distractors reduces the saliency of the distractors, and increases the relative saliency of the border region around the target by an even greater amount than familiarity of the target. Hence, these results are consistent with the psychophysical data showing that search for a novel target among familiar distractors is more efficient than search for a familiar target among novel distractors (Frith, 1974; Lubow & Kaplan, 1997; Reicher et al., 1976; Shen & Reingold, 2001; Wang et al., 1994; Wolfe, 2001). Furthermore, the difference in saliency between the results shown in columns 1 and 3 of Fig. 21 is consistent with data showing that search for a novel target is easier when the distractors are familiar rather then novel (Lubow & Kaplan, 1997).

4.3 Feature Search with Prior Exposure to Target or Distractor Features

While the psychophysical experiments described above have employed over-learned familiar stimuli, such as letters, the same effects can also be observed with a familiarity bias acquired only briefly before testing (Lubow & Kaplan, 1997). The model can also account for these results, if it is assumed that brief prior exposure to certain stimulus features will activate representations of those features at higher levels in the PC/BC model. In that case, these higher-level predictions will generate feedback to the PC/BC model of V1 which will affect processing of subsequent stimuli in an identical fashion to that shown in Fig. 21. Under this assumption, the model makes the reasonable prediction that visual input will remain constant (see section 3.10).

The difference in saliency between the results shown in the first two columns of Fig. 21 is therefore consistent with data showing that search is more efficient if human observers have prior knowledge about the nature of the target in a search array, either due to verbal instruction, or due to multiple trials with the same target being performed in blocks (Wolfe et al., 2003). This result is also consistent with the "priming of pop-out" effect, in which prior exposure to the features of the target increases the saliency of the target (Maljkovic & Nakayama, 1994). In addition, the difference between the results shown in the last two columns of Fig. 21 is consistent with the "distractor previewing effect", in which prior exposure to the features of the distractors improves search efficiency more than does prior exposure to the features of the target (Wan & Lleras, 2010).

4.4 Preview Search

The priming that occurs during a preview search experiment could be interpreted as an even more recent familiarity bias, one obtained during a trial rather than immediately or shortly prior to a trial (section 4.3), or in the distant past (section 4.2). The PC/BC model of V1 can simulate the preview effect without feedback from extrastriate area (section 3.9). In the model, increased saliency of the new items is due to the V1 prediction neurons taking time to accurately represent new items after onset. Hence, initially after onset the new items generate higher errors in comparison to the old items which the V1 prediction neurons have had time to represent more accurately. This effect would be increased by top-down predictions from higher-level processing stages. Specifically, during the preview the old items might be expected to activate higher-level perceptual representations encoding the location, features, or both, of the old items. Feedback from these neurons after the onset of the new items, would enhance the response of the V1 prediction neurons representing the old items, and hence, further reduce the error associated with these old items, enhancing the relative saliency of the new items.

4.5 Feature Search with Distractors in a Familiar Configuration

As described above, in the model a top-down prediction reduces the saliency of items consistent with that prediction. The effect is consistent with psychophysical experiments showing that search for a uniquely oriented line among heterogeneous distractors is easier when the distractors are tangent to a circle (Moraglia, 1989). In other words, the target becomes more salient when it does not conform to the circular shape formed by the other elements. The model would explain this by suggesting that the distractors organized in a circle would activate a higher-level representation of a circle which would in turn provide top-down excitation to the V1 prediction neurons representing those elements forming that circle. The error corresponding to these elements would then decrease, making the target (which is not part of the circle, and hence, not receiving top-down excitation) relatively more salient. In this account, the circle might be described as a familiar shape: one that is more easily predicted, and hence, less salient.

4.6 Feature Search with Contextual Guidance

Search efficiency is known to improve when the arrangement of distractors in a search array is repeated over many trials (Chun, 2002; Chun & Jiang, 1998). Top-down effects in the model can also explain this effect of contextual

guidance on visual search. If higher-level areas learn about the distractor array, then subsequent repetition of this array would activate the (recently learned) higher-level representation which would in turn provide feedback to V1. This feedback would increase the response of the prediction neurons at the learned locations and in turn reduce the error at those location. Increasing the relative salience of the target in a manner identical to that illustrated in Fig. 21, column 3.

4.7 Saliency of Incongruous Objects

Objects that are incongruous with the rest of the visual scene (like an octopus in a bathroom) are more likely to attract attention (Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood et al., 2008). The explanation offered by the PC/BC model is as follows. Prediction neurons at higher-levels in the cortical hierarchy, representing the brain's internal model of bathrooms, are strongly activated by the many parts of the scene consistent with this interpretation. Activity fed-back from these neurons will subsequently increase the response of V1 prediction neurons consistent with this interpretation, and hence, the error at these locations will be reduced. Lower-level representations that are inconsistent with the top-down interpretation will not receive feedback, and hence, the error at these locations is not decreased, making such locations relatively more salient than locations consistent with the high-level interpretation.

The combined effect of top-down predictions from extrastriate cortical areas is to increase the response of V1 prediction neurons encoding information consistent with the top-down expectation, and to reduce the response of error-detecting neurons encoding information consistent with the top-down expectation. The effect of the former is to make *consistent* information more conspicuous for cortical regions at subsequent stages along the processing hierarchy, and hence, make it more likely to be perceived. The effect of the latter is to make locations containing *inconsistent* information more salient, and hence, make these locations more likely to attract attention. These dual effects might explain the seemingly contradictory observations that congruent objects can be recognized more quickly than incongruent ones (Bar, 2004; Biederman et al., 1982; Boyce et al., 1989), but that incongruent objects are more likely to attract attention (Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood et al., 2008), and more generally the distinction between attended and expected stimuli (Summerfield & Egner, 2009). In the PC/BC model, saliency is distinct from perceptibility: activation in prediction neurons, which can be enhanced by top-down attention or expectation, is related to object recognition while activation in the error-detecting neurons, which can be enhanced by violations of expectation, is related to saliency.

4.8 Inhibition of Return in Serial Search Tasks

In order to efficiently perform a serial search task, the brain's attentional system has a preference to explore new locations in an image rather than re-fixate locations that have already been searched (Gilchrist & Harvey, 2000; Klein, 2000; Mirpour et al., 2009). This is believed to be due to the suppression of the saliency of previously visited locations, a mechanism referred to as " Inhibition of Return" (IoR; Gilchrist & Harvey, 2000; Itti & Koch, 2001; Klein, 2000; Mirpour et al., 2009; Posner & Cohen, 1984).

In the PC/BC model, attention to a specific spatial location would be expected to have similar effects as expectations generated by prior exposure to an image or features of an image, and to have similar effects to those caused by the familiarity of a shape or spatial configuration (as described in the preceding sections). Namely, attention would generate excitatory feedback (Spratling, 2008a) that would reduce the prediction error at, and saliency of, the attended location. By lowering saliency this increases the likelihood that attention will be attracted elsewhere. Furthermore, if it is assumed that fixating a location with the high resolution fovea, allows a more accurate representation of that location to be formed in extrastriate cortical areas, then this improved internal model will enable a more accurate reconstruction of the input at that location, reducing the saliency of that location, and hence making it less likely to attract attention in the future. This does not preclude the same location being fixated multiple times, since as the internal model is updated following further eye movements, old information may be overwritten or forgotten, making previously fixated locations salient again. Hence, the PC/BC model proposes that behavior consistent with IoR occurs as a side-effect of predictive coding.

5 Discussion

5.1 Comparison with Other Computational Models of Visual Saliency

In addition to hypothesizing that V1 is involved in calculating perceptual saliency, Zhaoping Li has proposed a specific model of how V1 might implement this saliency calculation (Li, 1999a,b, 2000, 2002; Zhaoping, 2003). In this model each location in the image is processed by a number of neurons selective for different orientations.

The response of each neuron is determined both by the feedforward activation from the image, and by lateral connections that enable the response to be modified by contextual influences. Excitatory lateral connections link neurons with spatially separate RFs but with orientation preferences that align to form part of a smooth contour. Inhibitory lateral connections link neurons with spatially separate RFs that are selective for approximately parallel orientations. This model is thus one specific example of a large class of models that postulate an "association field" of long-range lateral excitatory and inhibitory connections (*e.g.*, Ben-Shahar & Zucker, 2004a; Bosking et al., 1997; Dakin & Baruch, 2009; Field et al., 1993; Hess & Field, 1999; Hess et al., 2003; Kapadia et al., 1995, 2000; Mundhenk & Itti, 2005; Parent & Zucker, 1989; Yen & Finkel, 1998).

Filter-rectify-filter (FRF) models of texture segmentation have also been successfully applied to some of the tasks described in this article (*e.g.*, Baker & Mareschal, 2001; Bergen & Landy, 1991; Field et al., 1993; Landy & Bergen, 1991; Malik & Perona, 1990). Such models initially apply a number of filters to the image in order to generate separate maps of filter responses for a number of image properties. A nonlinear operation is then performed on the outputs of these filters, such as half-wave rectification or squaring, before a second set of filters are applied to find locations where the outputs of the first stage change. The V1 model of Zhaoping Li can be interpreted as a specific implementation of this general class of models, one in which the RFs of the orientation tuned neurons perform the first stage filtering, and the lateral connections perform the second stage of filtering (Zhaoping et al., 2009a).

"Saliency map" models of visual saliency have also been highly influential (*e.g.*, Bruce & Tsotsos, 2009; Itti & Koch, 2001; Koch & Ullman, 1985; Navalpakkam & Itti, 2006; Treisman & Souther, 1985; Treisman & Gelade, 1980; Wilder et al., 2011; Wolfe, 1994). These models propose that the stimulus is analyzed in parallel by a number of separate feature maps, within which the saliency of each location is determined by the degree to which the stimulus differs from its neighbors within that feature dimension. The outputs of these separate feature maps are then combined to determine the overall saliency at each location. Zhaoping Li's model is also closely related to saliency maps models, differing only in the following three respects (Koene & Zhaoping, 2007; Zhaoping et al., 2009b): (1) in assuming that the overall saliency is the max (rather than the sum) of the saliencies calculated across different feature dimensions; (2) in the identification of neural circuitry implementing the model, specifically that the saliency across individual feature dimensions are all calculated in V1 rather than in physically separate feature maps; and (3) that saliency may be calculated across specific feature combinations corresponding to the tuning properties of V1 cells.

The similarity between Zhaoping Li's model and FRF and saliency map models is not surprising since all these existing models implement the same fundamental computation in order to identify locations with high saliency. In all these models, salient locations are identified as those where the stimulus differs from its surroundings with respect to some feature or features. This mechanism can be thought of as a crude implementation of predictive coding: one where feature values at one location are predicted to be the same as feature values at neighboring locations, and salient locations are those where this prediction proves most erroneous. This is identical to the form of predictive coding that has been proposed to operate in the retina and which is believed to be implemented by the center-surround receptive field properties of retinal ganglion cells (Laughlin, 1990; Srinivasan et al., 1982). The form of predictive coding implemented in PC/BC is distinct in:

1. Proposing a more general and flexible model of the world: one that is defined by the receptive fields of the prediction neurons. These RFs could take any form and the model would equate salient locations with those that are most under-represented by the responses of the prediction neurons. However, for the purposes of this article the prediction neuron RFs have been defined as Gabor functions to be similar to V1 simple cell RFs. Even with this very simple model of the world, PC/BC successfully accounts for a very wide range of empirical results.

2. Maintaining a separate representation of the world. PC/BC employs distinct populations of neurons to represent predictions about the underlying causes of the sensory input, and to represent the error between the reconstruction of the input and the actual input. This enables the model to distinguish perception from salience (see section 4.7). Furthermore, this enables the errors to be used to iteratively update the predictions, and hence, to recursively improve the representation of the world encoded by the activity of the prediction neurons. In Zhaoping Li's model, and most implementations of saliency map models, the prediction errors overwrite the predictions.

The differences between the mechanism used by PC/BC to determine saliency, and the mechanism used by Zhaoping Li's model, saliency map models, and FRF models leads to several predictions that are distinct from those made by these previous models, as described in the next section.

A Bayesian surprise model of saliency has recently been proposed (Baldi & Itti, 2010; Itti & Baldi, 2009) that is distinct from the other models discussed above. In the Bayesian surprise model, saliency is related to the magnitude of the change in the brain's model of the world induced by the presentation of the data. In PC/BC,

the internal world model is represented by the response of the prediction neurons. Hence, in terms of PC/BC, Bayesian surprise would correspond to a measure of the change in the response of the prediction neurons from one iteration to the next. As a result, the Bayesian surprise model makes different predictions compared to PC/BC.

5.2 Novel Predictions of the PC/BC Model

Bayesian surprise proposes that when the model of the world provides a poor prediction of a stimulus, that stimulus will not be salient (Baldi & Itti, 2010). In contrast, PC/BC proposes the opposite; that an erroneous prediction will be salient in order to induce action or learning to improve the internal model.

As described in the previous section, all other models of visual saliency (Zhaoping Li's model of V1, FRF models, and saliency map models) presume that high saliency is correlated with a strong response from those neurons representing that stimulus. This mechanism necessarily confounds contrast and salience, and hence, high contrast stimuli will always be salient in such models. In the PC/BC model high salience is correlated with a strong response from neurons representing the error. It is thus possible for high contrast stimuli to be less salient than low contrast stimuli, as shown in Fig. 11.

Previous models assume that if V1 (or the range of feature maps available) lacks neurons with appropriate selectivities to encode the incoming information, the firing of the neurons in the model will be weak, and hence, the saliency low. In contrast, in the PC/BC model the lack of prediction neurons appropriate for representing the stimulus will result in a high salience due to an inaccurate reconstruction of the input.

Association field models, including Zhaoping Li's model, propose that contours are salient due to the mutual lateral excitation between neurons representing elements lying on a smooth contour. Hence, one would expect that if eye movements are driven by saliency then observers would fixate locations on the contour. In contrast, the PC/BC model proposes that it is the border between the elements forming the contour and the distractors that is salient. Observers would thus be expected to fixate distractor elements neighboring the contour at least as frequently as the contour itself. Similarly, during a search for a familiar target among novel distractors, the PC/BC model predicts the highest saliency, and hence, the most likely location for fixations would be on the distractors neighboring the target. In contrast, for a novel target among familiar distractors, the PC/BC model predicts that the target location will have the highest saliency, and hence, be the most likely target for saccades.

At the border between two textures made up of elements at 0^o and 90^o the PC/BC model predicts that the highest salience will occur on those elements adjacent and perpendicular to the border (see column 5 of Fig. 3b). In contrast, Zhaoping Li's model proposes that the highest saliency will occur on those elements adjacent and parallel to the border (Jingling & Zhaoping, 2008; Popple, 2003). Prior work to test this specific prediction of Zhaoping Li's model (Jingling & Zhaoping, 2008; Popple, 2003) has equated salience with the ability of an observer to perceive a change in a texture, or with the location at which an observer perceives the border. However, there is no reason to presume that either of these behavioral measures actually correlates with saliency.

Finally, the PC/BC model predicts that in an experiment on the contextual guidance in visual search, a familiar distractor array will speed the detection of the target even in situations where the target appears in different locations at each presentation.

5.3 Summary and Conclusions

PC/BC is an abstract, mathematical, model of cortical information processing (Spratling, 2011). As such it is intended to explore possible computational mechanisms, rather than the physiological mechanisms, underlying cortical function. Hence, the principal concern is with accurately simulating the behavior (of subjects or of single cells), and providing a computational explanation for this observed behavior, rather than describing the biological mechanisms through which it is achieved (Spratling, 2011). Nevertheless, the success of the model will be assessed below in terms of its ability to explain physiological data, as well as, its ability to explain psychophysical data.

5.3.1 Plausibility of the Model at the Implementation Level

It has previously been found that the computation preformed by the prediction neurons in the PC/BC model correlates very well with the behavior of orientation tuned cells in V1 (Spratling, 2010, 2011). There is also increasingly good evidence that the large-scale response of cortical populations, as measured using fMRI, is consistent with predictive coding (Egner et al., 2010). However, as discussed in several previous publication about PC/BC (Spratling, 2008b, 2010, 2011), while many aspects of the model are directly consistent with cortical anatomy and physiology, other aspects are not. Given the emphasis in this work on the behavior of the error-detecting neurons, it is interesting to speculate about how the function of these computational elements of the model might be implemented in the brain.

One possibility is that the structure of the model is fairly directly mapped onto cortical circuitry, and the error-detecting neurons represent a sub-population of cortical cells that has not been identified experimentally. The error-detecting neurons in the model have centre-surround RFs. While there is evidence for centre-surround cells in cortical layer IV of V1 in some species, there is no evidence for such cells in other species (Van Hooser, 2007). However, the error-detecting neurons make up a very small minority (one in 33) of the neurons in the PC/BC model of V1. If the model was extended to simulate other aspects of V1 behavior (such as color selectivity, disparity tuning, motion selectivity, *etc.*) this would require additional prediction neurons with a greater range of RF properties, and hence, error-detecting neurons would be predicted to make up an even smaller proportion of V1 cells. Furthermore, error-detecting neurons in the model are almost completely unselective to different stimuli (See Spratling, 2010, Supplemental Material, Fig.S6). Such rare and unselective cells could easily be over-looked in physiological experiments. Cortical cells with response properties that are influenced by context (*e.g.*, that exhibit surround suppression and modulation of firing rate by flankers), are often assumed to be the neural correlates of saliency computations and of texture segmentation. However, in the PC/BC model these contextual effects are accurately modeled by the behavior of the prediction neurons (Spratling, 2010).

A second possibility is that the very simple circuitry of the model does not map in a straight-forward way onto cortical circuitry. In other words, the complex circuitry of the cortex might achieve predictive coding in a very different manner from that suggested by a literal interpretation of PC/BC. In this case, the computation preformed by the error-detecting neurons in the model, might be performed by the combined action of many different classes of cortical cell, including different sub-populations of inhibitory inter-neuron. A third possibility is that the error-detecting neurons do not reside in cortex, but instead the prediction values might be read out from V1 and the error calculated subcortically by local comparison of these predictions and the retinal input. In this case, the error-detecting neurons might reside in the superior colliculus (SC) or pulvinar.

The PC/BC model proposes that prediction error influences covert and overt attention. This influence over behavior might be achieved via connections to the SC and pulvinar (Shipp, 2004). This is the case whether those connections transmit error signals calculated cortically, or if they transmit predictions to allow error to be calculated locally. However, for the sake of convenience, the rest of this paragraph assumes the former. Many different areas of the cortex which process visual information send connections to SC and pulvinar (Shipp, 2004). Hence, each stage in the cortical visual processing hierarchy might calculate error signals which are combined in SC. Furthermore, V1 represents the top of a cortical hierarchy of feedback connections (Casagrande et al., 2005), and the error calculated in V1 could be influenced by a cascade of feedback, thereby enabling higher cortical areas to affect the output sent to SC from V1. These two options are not mutually exclusive and either could account for the saliency of certain higher-level image properties, such as an element with a unique apparent 3D depth/shape (Enns & Rensink, 1990; Nakayama & Silverman, 1986; Smith et al., 2007) or an incongruous object in a natural scene (Loftus & Mackworth, 1978; Underwood & Foulsham, 2006; Underwood et al., 2008)

5.3.2 Plausibility of the Model at the Behavioral and Computational Levels

The PC/BC model succeeds in simulating a very large range of psychophysical data concerning the saliency of borders between orientation and luminance defined textures, the range of saliencies for unique orientations, shapes and contrasts in search arrays, asymmetry in feature search, preview search, the saliency of abrupt onsets, the effects of element novelty and familiarity in feature search, contextual guidance in feature search, priming of pop-out, the distractor previewing effect, change blindness, and the saliency of contours embedded in random texture.

Importantly, the model also successfully simulates experiments designed to demonstrate that V1 is involved in the calculation of saliency (Zhaoping & May, 2007). Hence, the model provides an alternative mechanisms by which a bottom-up V1 saliency map could be implemented. However, the performance of the PC/BC model goes beyond that of the only other model of the V1 saliency map hypothesis (Li, 1999a,b, 2000, 2002; Zhaoping, 2003) in several respects:

- Modeling the effects of luminance on salience.
- Modeling the effects of spacing on salience.
- Modeling the effects of abrupt onsets.
- Modeling the preview benefit.
- Modeling change blindness.
- Modeling the effects of familiarity and top-down knowledge on salience.

- Suggesting a mechanism for IoR.
- Modeling a wide range of properties of orientation-tuned cells in V1 (Spratling, 2010, 2011).
- Modeling a wide range of neurophysiological data on the effects of endogenous attention (Spratling, 2008a).

Despite the success of the current model there are clear ways in which it could be improved. One is by explicitly modeling the behavior of extrastriate areas rather than simulating their effects via an additional input to the model of V1. This could be done by modeling each extrastriate area as a processing stage in a PC/BC hierarchy. The challenge would be to define appropriate synaptic weights for the neurons in these higher levels. Another obvious improvement that could be made to the current model is to incorporate mechanisms for processing color information. This would require the inclusion of V1 prediction neurons with RFs selective for color. This would be especially useful since many visual search tasks employ colored stimuli. Another problem is that the model is somewhat over sensitive to the exact position and orientation of stimuli. This seems to be the result of modeling V1 RFs using a very regular and uniform set of Gabor functions. There is thus a very limited range of RF shapes and sizes. This might easily be fixed by defining a wider range of RFs, however, this has not been done here in order that this model of V1 is identical to that used previously to simulate other aspects of V1 behavior (Spratling, 2010, 2011).

The PC/BC model provides a single computational explanation for a very wide range of saliency effects. The model proposes that high saliency occurs at locations in an image where the brain's internal representation of the world (which is encoded by the activity of neurons in V1 and multiple extrastriate cortical regions) is least able to account for the sensory data. Given that salient locations are known to attract eye movements or be targets for covert attention, PC/BC further proposes that locations at which the brain's internal model of the world are least accurate will give rise to action that can reduce this error. Specifically, directing the high acuity fovea to a location where the model is poor will provide information at a finer resolution which can be used to update and increase the fidelity of the internal model. Furthermore, if high prediction error captures endogenous attention then this will increase the response of prediction neurons representing that location, and so increase the likelihood that this information propagates up the cortical hierarchy to be represented by other prediction neurons and so become included in the internal model. In addition to short-term effects on the currently active representation of the world, saliency might also have long-term effects by enabling learning to adapt the cortical representation where this representation is poorest (Bubic et al., 2010). The role of prediction errors in driving action to reduce those errors, in both the short and long term, is consistent with the free-energy principle (Friston, 2010; Friston et al., 2010).

Acknowledgments

This work was funded by the Engineering and Physical Sciences Research Council grant number EP/D062225/1.

References

Attneave, F. (1954). Informational aspects of visual perception. *Psychol. Rev.*, *61*, 183–93.

Baker, C. L., & Mareschal, I. (2001). Processing of second-order stimuli in the visual cortex. *Prog. Brain Res.*, *134*, 171–91.

Baldi, P., & Itti, L. (2010). Of bits and wows: A bayesian theory of surprise with applications to attention. *Neural Netw.*, *23*, 649–66.

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.*, *5*, 617–29.

Barlow, H. (2001). Redundancy reduction revisited. *Network*, *12*, 241–53.

Barlow, H. B. (1994). What is the computational goal of the neocortex? In C. Koch, & J. L. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain* chapter 1. (pp. 1–22). Cambridge, MA: MIT Press.

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Res.*, *36*, 1439–66.

Ben-Shahar, O., & Zucker, S. W. (2004a). Geometrical computations explain projection patterns of long range horizontal connections in visual cortex. *Neural Comput.*, *16*, 445–76.

Ben-Shahar, O., & Zucker, S. W. (2004b). Sensitivity to curvatures revealed in orientation-based texture segmentation. *Vision Res.*, *44*, 257–77.

Bergen, J. R., & Landy, M. S. (1991). Computational modelling of visual texture segregation. In M. S. Landy, & M. J. Anthony (Eds.), *Computational modelling of visual processing*. Cambridge, MA: MIT Press.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.*, *14*, 143–77.

Bosking, W., Zhang, Y. B. S., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in the tree shrew striate cortex. *J. Neurosci.*, *17*, 2112–27.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object recognition. *J. of Expt. Psyc.: Human Percept. & Perform.*, *15*, 556–66.

Bruce, N. D. B., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. *J. Vis.*, *9*, 1–24.

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.*, *4*, 1–15.

Carrasco, M., McLean, T. L., Katz, S. M., & Frieder, K. S. (1998). Feature asymmetries in visual search: effects of display duration, target eccentricity, orientation and spatial frequency. *Vision Res.*, *38*, 347–74.

Casagrande, V. A., Sáry, G., Royal, D., & Ruiz, O. (2005). On the impact of attention and motor planning on the lateral geniculate nucleus. In V. A. Casagrande, R. W. Guillery, & S. M. Sherman (Eds.), *Cortical Function: a View from the Thalamus* (pp. 11–29). Elsevier volume 149 of *Progress in Brain Research*.

Chun, M. M. (2002). Contextual cueing of visual attention. *Trends Cogn. Sci.*, *4*, 170–8.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.*, *36*, 28–71.

Dakin, S. C., & Baruch, N. J. (2009). Context influences contour integration. *J. Vis.*, *9*, 1–13.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res.*, *20*, 847–56.

Daugman, J. G. (1988). Complete discrete 2-D Gabor transformations by neural networks for image analysis and compression. *IEEE Trans. Acoust.*, *36*, 1169–79.

Donk, M. (2005). Prioritizing selection of new elements: on the time course of the preview effect. *Vis. Cogn.*, *17*, 1373–85.

Donk, M. (2006). The preview benefit: visual marking, feature-based inhibition, temporal segregation, or onset capture? *Vis. Cogn.*, *14*, 736–48.

Donk, M., & Theeuwes, J. (2001). Visual marking beside the mark: Prioritizing selection by abrupt onsets. *Percept. Psychophys.*, *63*, 891–900.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychol. Rev.*, *96*, 433–58.

D'Zmura, M. (1991). colour in visual search. *Vision Res.*, *31*, 951–66.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.*, *30*, 16601–8.

Enns, J. T., & Rensink, R. A. (1990). Sensitivity to three-dimensional orientation in visual search. *Psychol. Sci.*, *1*, 323–6.

Fahle, M. (1991). Parallel perception of vernier offests, curvature, and chevrons in humans. *Vision Res.*, *31*, 2149–84.

Field, D., Hayes, A., & Hess, R. (1993). Contour integration in the human visual system: evidence for a local 'association' field. *Vision Res.*, *33*, 173–93.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, *4*, 2379–394.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Comput.*, *6*, 559–601.

Foster, D. H., & Ward, P. A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proc. R. Soc. Lond., B, Biol. Sci.*, *243*, 75–81.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.*, *11*, 127–38.

Friston, K. J. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, *360*, 815–36.

Friston, K. J. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.*, *13*, 293–301.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.*, *102*, 227–60.

Frith, U. (1974). A curious effect with reversed letters explained by a theory of schema. *Percept. Psychophys.*, *16*, 113–6.

Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. *Curr. Biol.*, *10*, 1209–12.

Giora, E., & Casco, C. (2007). Region- and edge-based configurational effects in texture segmentation. *Vision Res.*, *47*, 879–86.

Hess, R., & Field, D. (1999). Integration of contours: new insight. *Trends Cogn. Sci.*, *3*, 480–86.

Hess, R., Hayes, A., & Field, D. J. (2003). Contour integration and cortical processing. *J. Physiol. Paris*, *97*, 105–19.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Comput.*, *14*, 1711–1800.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The wake-sleep algorithm for unsupervised neural networks. *Science*, *268*, 1158–61.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.*, *5*, 1457–69.

Hoyer, P. O., & Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems*, *11*, 191–210.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Res.*, *49*, 1295–1306. Visual Attention: Psychophysics, electrophysiology and neuroimaging.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.*, *2*, 194–202.

Jingling, L., & Zhaoping, L. (2008). Change detection is easier tat texture border bars when they are parallel to the border: evidence for V1 mechanisms of bottom-up salience. *Perception*, *37*, 197–206.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, *58*, 1233–58.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Percept. Psychophys.*, *43*, 346–54.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*, 843–856.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *J. Neurophysiol.*, *84*, 2048–62.

Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *J. Cogn. Neurosci.*, *13*, 90–101.

Klein, R. M. (2000). Inhibition of return. *Trends Cogn. Sci.*, *4*, 138–47.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.*, *4*, 219–27.

Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *J. Vis.*, *7*, 1–14.

Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Res.*, *31*, 67991.

Laughlin, S. (1990). Coding efficiency and visual processing. In C. Blakemore (Ed.), *Vision: Coding and Efficiency* chapter 2. (pp. 25–31). Cambridge University Press.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–91.

Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, *18*, 959–71.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, *20*, 1434–48.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U.S.A.*, *99*, 8378–83.

Li, W., & Gilbert, C. D. (2002). Global contour saliency and local colinear interactions. *J. Neurophysiol.*, *88*, 2846–56.

Li, Z. (1999a). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc. Natl. Acad. Sci. U.S.A.*, *96*, 10530–5.

Li, Z. (1999b). Visual segmentation by contextual influences via intracortical interactions in primary visual cortex. *Network*, *10*, 187–212.

Li, Z. (2000). Pre-attentive segmentation in the primary visual cortex. *Spatial Vis.*, *13*, 25–50.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends Cogn. Sci.*, *6*, 9–16.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *J. of Expt. Psyc.: Human Percept. & Perform.*, *4*, 565–72.

Lubow, R. E., & Kaplan, O. (1997). Visual search as a function of type of prior experience with target and distractor. *J. of Expt. Psyc.: Human Percept. & Perform.*, *23*, 14–24.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, *7*, 923–32.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. role of features. *Mem. Cognit.*, *22*, 657–72.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, *70*, 1297–1300.

Mirpour, K., Arcizet, F., Ong, W. S., & Bisley, J. W. (2009). Been there, seen that: A neural mechanism for performing efficient visual search. *J. Neurophysiol.*, *102*, 3481–91.

Moraglia, G. (1989). Display organization and the detection of horizontal line segments. *Percept. Psychophys.*, *45*, 265–72.

Mounts, J. R. W., & Tomaselli, R. G. (2005). Competition for representation is mediated by relative attentional

salience. *Acta Psychol.*, *118*, 261–75.
Mumford, D. (1992). On the computational architecture of the neocortex II: the role of cortico-cortical loops. *Biol. Cybern.*, *66*, 241–51.
Mundhenk, T. N., & Itti, L. (2005). Computational modeling and exploration of contour integration for visual saliency. *Biol. Cybern.*, *93*, 188–212.
Nagy, A. L., & Sanchez, R. R. (1992). Chromaticity and luminance as coding dimensions in visual search. *Human Factors*, *34*, 601–14.
Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*, 264–5.
Navalpakkam, V., & Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimizing detection speed. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (pp. 2049–56).
Nothdurft, H.-C. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Res.*, *25*, 1957–68.
Nothdurft, H.-C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Res.*, *31*, 1073–78.
Nothdurft, H.-C. (1992). Feature analysis and the role of similarity in preattentive vision. *Percept. Psychophys.*, *52*, 355–75.
Nothdurft, H.-C. (2000). Salience from feature contrast: variations with texture density. *Vision Res.*, *40*, 3181–200.
Nothdurft, H.-C. (2006). Salience-controlled visual search: are the brightest and the least bright targets found by different processes? *Vis. Cogn.*, *13*, 700–32.
Olivers, C. N. L., Humphreys, G. W., & Braithwaite, J. J. (2006). The preview search task: evidence for visual marking. *Vis. Cogn.*, *14*, 716–35.
Olshausen, B. A., & Field, D. J. (1996a). Emergence of simple-cell receptive properties by learning sparse code for natural images. *Nature*, *381*, 607–9.
Olshausen, B. A., & Field, D. J. (1996b). Natural image statistics and efficient coding. *Network*, *7*, 333–9.
Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.*, *37*, 3311–25.
Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding V1? *Neural Comput.*, *17*, 1665–99.
Olson, R. K., & Attneave, F. (1970). What variables produce similarity grouping? *Am. J. Psychol.*, *83*, 1–21.
Parent, P., & Zucker, S. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, *11*, 823–39.
Pashler, H., Dobkins, K., & Huang, L. Q. (2004). Is contrast just another feature for visual selective attention? *Vision Res.*, *44*, 1403–10.
Popple, A. V. (2003). Context effects on texture border localization bias. *Vision Res.*, *43*, 739–43.
Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma, & D. Bouwhuis (Eds.), *Attention and Performance X* (pp. 531–56). Hillsdale, NJ: Lawrence Erlbaum Associates.
Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.*, *2*, 79–87.
Reicher, G. M., Snyder, C. R. R., & Richards, J. T. (1976). Familiarity of background characters in visual scanning. *J. of Expt. Psyc.: Human Percept. & Perform.*, *2*, 522–30.
Rensink, R., O'Regan, J., & Clark, J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychol. Sci.*, *38*, 368–73.
Rensink, R., O'Regan, J., & Clark, J. (2000). On the failure to detect changes in scenes across brief interruptions. *Vis. Cogn.*, *37*, 127–45.
Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, *2*, 1019–25.
Saiki, J., Koike, T., Takahashi, K., & Inoue, T. (2005). Visual search asymmetry with uncertain targets. *J. of Expt. Psyc.: Human Percept. & Perform.*, *31*, 1274–87.
Serre, T., Oliva, A., & Poggio, T. (2007a). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.*, *104*, 6424–9.
Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007b). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.*, *29*, 411–26.
Shen, J., & Reingold, E. M. (2001). Visual search asymmetry: The influence of stimulus familiarity and low-level features. *Percept. Psychophys.*, *63*, 464–75.
Shipp, S. (2004). The brain circuitry of attention. *Trends Cogn. Sci.*, *8*, 223–30.
Smith, M. A., Kelly, R. C., & Lee, T. S. (2007). Dynamics of response to perceptual pop-out stimuli in macaque V1. *J. Neurophysiol.*, *98*, 3436–49.
Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual selective attention. *Vision Res.*, *48*, 1391–408.

Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Front. Comput. Neurosci.*, *2*, 1–8.
Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *J. Neurosci.*, *30*, 3531–43.
Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. *Vision Res.*, *51*, 563–76.
Spratling, M. W. (in press). Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function. *Neural Comput.*, *in press*.
Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proc. R. Soc. Lond., B, Biol. Sci.*, *216*, 427–59.
Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn. Sci.*, *13*, 403–9.
Theeuwes, J., Kramer, A. F., & Atchley, P. (1998). Visual marking of old objects. *Psychon. Bull. Rev.*, *5*, 130–4.
Thorpe, S. J., Guyonneau, R., Guilbaud, N., Allegraud, J. M., & VanRullen, R. (2004). Spikenet: Real-time visual processing with one spike per neuron. *Neurocomputing*, *58-60*, 857–64.
Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychol. Rev.*, *95*, 15–48.
Treisman, A., & Souther, S. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *J. of Expt. Psyc.: General*, *114*, 285–310.
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.*, *12*, 97–136.
Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Q. J. Exp. Psychol.*, *18*, 1931–49.
Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? evidence from eye movements during the inspection of real-world scenes. *Conscious. Cogn.*, *17*, 159–70.
Van Hooser, S. D. (2007). Similarity and diversity in visual cortex: Is there a unifying theory of cortical computation? *Neuroscientist*, *13*, 639–56.
VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? is it a plane? ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*, 655–68.
Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Prog. Neurobiol.*, *51*, 167–94.
Wan, X., & Lleras, A. (2010). The effect of feature discriminability on the intertrial inhibition of focused attention. *Vis. Cogn.*, *18*, 920–44.
Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Percept. Psychophys.*, *56*, 495–500.
Watson, D., & Humphreys, G. W. (1997). Visual marking: prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychol. Rev.*, *104*, 90–122.
Watt, R., Ledgeway, T., & Dakin, S. C. (2008). Families of models for gabor paths demonstrate the importance of spatial adjacency. *J. Vis.*, *8*, 1–19.
Wilder, M. H., Mozer, M. C., & Wickens, C. D. (2011). An integrative, experience-based theory of attentional control. *J. Vis.*, *11*.
Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychon. Bull. Rev.*, *1*, 202–38.
Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). London, UK: University College London Press.
Wolfe, J. M. (2001). Asymmetries in visual search: an introduction. *Percept. Psychophys.*, *63*, 381–9.
Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends Cogn. Sci.*, *7*, 70–6.
Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *J. of Expt. Psyc.: Human Percept. & Perform.*, *29*, 483–502.
Wolfe, J. M., & Friedman-Hill, S. R. (1992). On the role of symmetry in visual search. *Psychol. Sci.*, *3*, 194–8.
Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *J. of Expt. Psyc.: Human Percept. & Perform.*, *18*, 34–49.
Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.*, *5*, 1–7.
Wolfson, S. S., & Landy, M. S. (1995). Discrimination of orientation-defined texture edges. *Vision Res.*, *35*, 2863–77.
Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *J. of*

Expt. Psyc.: Human Percept. & Perform., *10*, 601–21.

Yen, S., & Finkel, L. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Res.*, *38*, 719–41.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.*, *10*, 301–8.

Zhaoping, L. (2003). V1 mechanisms and some figure-ground and border effects. *J. Physiol. Paris*, *97*, 503–15.

Zhaoping, L. (2008). Attention capture by eye of origin singletons even without awareness – a hallmark of a bottom-up saliency map in the primary visual cortex. *J. Vis.*, *8*, 1–18.

Zhaoping, L., Guyader, N., & Lewis, A. (2009a). Relative contributions of 2D and 3D cues in a texture segmentation task, implications for the roles of striate and extrastriate cortex in attentional selection. *J. Vis.*, *9*, 1–22.

Zhaoping, L., & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Comput. Biol.*, *3*, e62.

Zhaoping, L., May, K. A., & Koene, A. (2009b). Some fingerprints of V1 mechanisms in the bottom up saliency for visual selection. In D. Heinke, & E. Mavritsaki (Eds.), *Computational modelling in behavioural neuroscience: closing the gap between neurophysiology and behaviour* chapter 7. (pp. 137–64). London, UK: Psychology Press.